INTERPRETATIONS OF PROBABILITY

Andrei Khrennikov

///VSP///

Utrecht • Boston, 2003

VSP
an imprint of Brill Academic Publishers
P.O. Box 346
3700 AH Zeist
The Netherlands

Tel: +31 30 692 5790
Fax: +31 30 693 2081
vsppub@compuserve.com
www.vsppub.com
www.brill.nl

First published in 1999
This edition 2003

ISBN 90-6764-310-6

Interpretations of Probability

by
Andrei Khrennikov
Växjö University, Sweden

June 10, 2003

I dedicate this book to Olga

3

Introduction 5

I Foundations of Probability Theory 13

1. A few words about measures 13
2. Classical and Ensemble Definitions of Probability 14
3. Frequency Theory of Probability 17
4. Kolmogorov's Measure-theoretical Approach 22
5. Kolmogorov's Ideas on Probability 25
6. Measure-theoretical Approach and Interpretations of Probability 32
7. Subjective (Bayesian) Probability Theory 39
8. Foundations of Randomness 41
9. Operation of Combining of Collectives 48
10. Independence of Collectives 51
11. Frequency and Measure-theoretical Viewpoints on Independence 53
12. Generalization of the Operation of Combining 54
13. Comparative Probability 55

II Quantum Probabilities 59

1. Classical and Quantum Probability Rules 59
2. Physical Interpretations of the Wave Function 64
3. 'Contradiction' between Quantum and Classical Probability Calculi 67
4. Probabilities with respect to Objective Conditions 77
5. Einstein–Podolsky–Rosen Paradox: Probability, Reality, Locality 83
6. Bell's Inequality for Probabilities 87
7. Bell's Mystification 91
8. Bell's Inequality for Covariations 94
9. Hidden Variables and Bell's Inequality 95

III Negative Probabilities 109

1. The Origin of Negative Probabilities in the Ensemble and Frequency Theories 109
2. Signed 'Probabilistic' Measures and Einstein–Podolsky–Rosen Paradox 120
3. Wiegener Phase-Space Distribution and Negative Probability . 128
4. Dirac's World with Negative Probabilities 135
5. Negative Probabilities and Localization 138

IV p-Adic Probability Theory **141**

1. Non-Archimedean Number Systems; p-Adic Numbers 143
2. Frequency Probability Theory 149
3. Ensemble Probability 152
4. Measures 162
5. p-Adic Probability Space 168

V Information and Probability **171**

1. Information Reality 171
2. Dynamics on Information Spaces 173
3. Information Velocity, Acceleration, Mass and Force, Newton's Laws 177
4. Mathematical 'Pathologies' 179
5. Quantum Mechanics on Information Spaces 180
6. The Pilot Wave Theory for Cognitive Systems 185
7. Subjective Probability as Probability with respect to the Statistical Ensemble of Human Ideas 189
8. Bayes' Theorem and Freud's Psychoanalysis 190

VI Tests for Randomness for p-adic Probability Theory **195**

1. p-Adic Probability Measures on the Space of Binary Sequences 196
2. Some Technical p-Adic Results 197
3. p-Adic Tests for Randomness 198
4. Limit Theorems 203
5. Recursive Enumeration of the Set of p-Adic Tests 205
6. No p-Adic Universal Test 208
9. Randomness of Infinite Sequences 209

Bibliography **215**

Index **225**

Introduction

The modern axiomatics of probability theory was created by Andrei Nikolaevich Kolmogorov in 1933. This axiomatics is based on the measure-theoretical approach to probability. The main advantage of the Kolmogorov probabilistic formalism is the high level of abstractness. The use of abstract probability spaces gives the possibility to develop general probabilistic calculus in that structures of concrete spaces of elementary events do not play any role. The Kolmogorov probabilistic formalism was successfully developed in many direction. This formalism is now the mathematical basis for numerous probabilistic models in physics, technique, biology, finance. Everything is perfect in the landscape of Kolmogorov probability theory except for one dark cloud obscuring the horizon.

This cloud is the probability foundation of quantum mechanics. This cloud was generated by A. Einstein, B. Podolsky and N. Rosen (EPR) in 1935 (just two years after creation of the axiomatics of probability theory). EPR started the discussion on *completeness* of quantum mechanics (the possibility to describe whole physical reality by the formalism of quantum mechanics). In fact, the problem of completeness has the close connection with foundations of probability theory. EPR proposed some arguments that can be interpreted as the evidence of incompleteness of quantum mechanics. The following discussion on the EPR arguments demonstrated that quantum mechanics has quite marshy foundations. Often the EPR arguments are even considered as the paradox in foundations of quantum mechanics. During following thirty years dark quantum cloud obscuring the landscape of the Kolmogorov probability theory was rather small. The probabilistic roots of the EPR paradox were not so evident. Nobody tried to connect the paradox in foundations of quantum mechanics with foundations of probability theory. The first attempt to provide the probabilistic representation of the EPR considerations was done by J. Bell who found in 1964 famous Bell's inequality

for covariations of physical observables involved in the EPR experiment. The black quantum cloud became quite large. Even in 1964 it is not only blotted out the sun of the Kolmogorov landscape, but it was gathering to obscure the beautiful idea of unique and general probability theory. However, nothing occurred in 1964. Moreover, nothing occurred in the following thirty years. And it seems to be that nothing is gathering to occur with unique and general Kolmogorov probability theory.

The great Kolmogorov probability community is still working in the standard measure-theoretical formalism. They do not pay attention to quantum clouds. On the other hand, the majority of the physical community observes this cloud. However, physicists do not understand the hidden probabilistic structure of this cloud. Some of them support the idea of the *death of reality.* They think that it is impossible to use realism in quantum considerations. And if there is no reality at all, they do not afraid this non-real cloud. Other physicists support the idea of *nonlocality.* They think that physical reality is nonlocal. Thus by doing some measurement for a quantum system in Moscow we change the quantum state of the correlated quantum system which is located in Vladivostok. The adherents of nonlocality also do not observe the black cloud: this cloud is distributed everywhere and, hence, such a cloud could not induce storm.

There are many reasons for this strange situation. One of them is purely psychological. Mathematicians are not interested in quantum physics (mainly because they still do not know quantum theory). Physicists are not interested in foundations of probability theory (mainly because they know not so much even about the standard Kolmogorov measure-theoretical approach). In principle, even J. Bell in 1964 could pay attention that Bell's inequality is connected not only with such properties of physical observables as realism and locality, but also with the way of the probabilistic description. However, this was not done[1]. Bell's inequality was not considered as a sign for reconsideration of the foundations of probability theory. In the opposite to geometry probability theory was not transformed in an elastic formalism containing numerous probabilistic models which can be used for descriptions of different physical phenomena. Probability theory is still a rigid structure. This structure can be compared with the rigid Euclidean cub. Attempts to use the unique Kolmogorov model for describing all physical phenomena can

[1]I had numerous discussions with scientists worked with J. Bell. Unfortunetely the general opinion is that J. Bell had never been interested in probability theory.

be compared with attempts to represent all geometrical models by Euclidean cubs. However, geometric reality is not restricted to reality of cubs as well as probabilistic reality is not restricted to reality of Kolmogorov probability spaces.

In this book we demonstrate that 'pathological behaviour' of 'quantum probabilities is a consequence of the use of Kolmogorov's approach. The high level of abstractness does not give the possibility to control connection between probabilities and *statistical ensembles* or *random sequences* (collectives). Formal manipulations with abstract Kolmogorov probabilities produce such monsters as Bell's inequality (in fact, this idea was already discussed by L. de Broglie and later improved by G. Lochak).

First attempts to reduce the EPR paradox to the use of one concrete probability model, namely, the Kolmogorov model, were the works of L. Accardi and I. Pitowsky. L. Accardy proposed rather formal non-Kolmogorovean model which did *not contain Bayes' formula.* I. Pitowsky proposed to consider events which are described by *nonmeasurable sets.* The latter formalism was strongly improved by S. Gudder who developed the theory of *probability manifolds.* The main disadvantage of all these models is that they have even higher level of abstractness than Kolmogorov's model. Therefore they provide merely a new description of quantum phenomena. They could not explain probabilistic roots of quantum behaviour. The same can be said about so called *quantum probabilities.* Of course, quantum probability calculus gives the useful and convenient description of quantum phenomena. However, quantum probability has no direct connection with probability. This is just rather speculative use of the word 'probability' in some formal mathematical constructions.

In this book we *explain* quantum probabilistic behaviour by using two basic interpretations of probability: *ensemble* and *frequency.* We demonstrate that (despite of the common opinion) the ensemble and frequency probability models are not in general equivalent to Kolmogorov's model.

The frequency model is von Mises' theory of collectives (random sequences), 1919. In this model probabilities are defined as limits of relative frequencies, $\nu_N = n/N$, in a collective x. This model was intensively studied in probability theory to find a reasonable definition of a random sequence (Kolmogorov algorithmic complexity and Martin-Löf theory of tests for randomness). It is the common opinion that frequency probabilities can be reduced to Kolmogorov probabilities. We explain (using the original arguments of R. von Mises) that such a viewpoint is totally wrong. In particular, the

law of large numbers does not describe the statistical stabilization of relative frequencies, $\nu_N = n/N$, in a collective x. It will be shown that the frequency probability model has many features which extremely differ from features of Kolmogorov's model. The most important feature of frequency probability is dependence of probabilities on a collectives x. In particular, the careful control of such a dependence gives the possibility to eliminate Bell's inequality from considerations. The frequency model also differs from Kolmogorov's model in the approach to conditional probabilities and independence. Such a difference also plays the important role in quantum considerations.

The ensemble model, as well as the frequency model, is one of basic 'pre-Kolmogorov' models. For example, the well known Bernoulli theorem is , in fact, a theorem for ensemble probabilities. It is commonly supposed that the definition of ensemble probability (as a proportion in an ensemble) is just a particular case of Kolmogorov's measure-theoretical definition. It is not right. The ensemble probability model cannot be reduced to Kolmogorov's one. The most important feature of ensemble probabilities is dependence on an ensemble. In particular,the careful control of such a dependence gives the possibility to eliminate Bell's inequality from considerations. It is impossible to use Kolmogorov's measure-theoretical approach for describing ensemble probabilities for infinite ensembles. In fact, Kolmogorov's measures are not proportional distributions of properties of ensembles of physical systems. These are measures on ensembles of all possible sequences of results of measurements. To obtain the adequate mathematical description of ensemble (proportional) probabilities for infinite ensembles (and quantum states describe such ideal ensembles), we have to leave the domain of Kolmogorov's probability model and, moreover, the domain of real analysis. We have to use number systems which contain *actual infinities.* In this book we use systems of so called p-adic numbers $\mathbf{Q}_p$ (where $p > 1$ are prime numbers) for the description of some infinite statistical ensembles ($\mathbf{Q}_p$ contains actual infinities).

The origin of p-adic ensemble probabilities can be illustrated by the following example. Let S be an infinite ensemble of balls. Each ball has some colour $c \in C = \{0, 1, 2, ..., k, ...\}$ (countable system of colours). The S has the following colour structure: there are $n_k = 2^k$ balls with the colour $k \in C$ in S. The 'volume' $N = |S|$ of S can be easily found:

$$N = \sum_{k=0}^{\infty} n_k = \sum_{k=0}^{\infty} 2^k.$$

Of course, this series diverges in the field of real numbers $\mathbf{R}$. But it converges in the field of 2-adic numbers $\mathbf{Q}_2$. The sum of this series $\mathbf{Q}_2$ can be found by using ordinary formula for the sum of infinite geometric progression (because $2^k \to 0, k \to \infty$, in $\mathbf{Q}_2$) :

$$N = \sum_{k=0}^{\infty} 2^k = \frac{1}{1-2} = -1.$$

We can now find the proportion of balls with the colour $k \in C$ in the ensemble S :

$$\mathbf{P}_S(k) = \frac{n_k}{N} = -2^k.$$

We remark that, as $n_k = 2^k$ is a finite number and $N = -1$ is an infinite number, the probability $\mathbf{P}_S(k)$ is *infinitely small probability.* And such a probability is represented by a negative number. This approach induces the rigorous mathematical theory of *negative probabilities.*

In fact, negative probabilities is other cloud obscuring the Kolmogorov probability landscape. Negative probabilities (which could not be justified by Kolmogorov's model) arise with the strange regularity in practically all quantum models. The most famous are Wiegner distribution on the phase space and Dirac's negative probability distributions in the formalism of relativistic quantization. Such 'probability distributions' are considered as monsters of quantum theory. For example, physicists always underline that Wiegner distribution is not really a probability distribution. At the same time they continue to use it for describing probabilistic phenomena. In this book negative probabilities (in particular, the Wiegner distribution) are realized as probabilities with respect to infinitely large statistical ensembles. In many physical models these probability have the interpretation of infinitely small probabilities. Negative probabilities can also be obtained in the frequency approach as limits of relative frequencies, $\nu_N = n/N$, with respect to some topology on the set rational numbers $\mathbf{Q}$ which differs from the standard real topology (and frequencies $\nu_N = n/N$ always belong to $\mathbf{Q}$). For example, in the p-adic topology the probability $\mathbf{P} = -1$ can be obtained as the limit of frequencies:

$$P = -1 = \lim \nu_N.$$

Typically in the frequency approach the presence of negative probabilities is the exhibition of the violation of the principle of the statistical stabilization for relative frequencies with respect to the real topology. Negative frequency

probabilities can also appear via, for example, p-adic splitting of conventional probability $\mathbf{P} = 0$. In the latter case negative probabilities can be again interpreted as infinitely small probabilities.

In a part of book we investigate connection between information and probability. A model of purely information physical reality is developed: here basic objects are information objects (so called transformers of information), physical processes are purely information processes and statistics is information statistics. We investigate models of classical and quantum mechanics (in particular, the pilot wave theory) on information spaces. Such information mechanics can be used for describing cognitive, social, psychological and even anomalous phenomena. Subjective probability is considered from the information viewpoint (as the ensemble probability for information ensembles; in particular, ensembles of human ideas). Such an interpretation of subjective probabilities is used for studying some psychological phenomena. We try to explain Freud's psychoanalysis by considering ensembles of conscious and unconscious ideas and roles of these two classes of ideas in Bayes' formula for conditional probabilities.

The last chapter of the book has purely mathematical character. Here we develop a p-adic analogue of the Martin-Löf theory of tests for randomness (to find a p-adic analogue of a random sequence). p-adic theory strongly differs from the real one. There is no universal test for randomness (at least for the uniform p-adic probability distribution). In this sense the p-adic theory of randomness is similar to Schnorr's theory for Kolmogorov probabilities. We also obtained a large class of limit theorems for p-adic probabilities. In particular, these limit theorems can be applied to negative probabilities. We remark that the first limit theorem for negative probabilities was proved by M. Barnett in 1944.

Main consequences of the book:

1. Kolmogorov's probability theory (measure-theoretical approach) is just one of many probability model.

2. Two fundamental interpretations of probability, namely, the *ensemble and frequency interpretations,* can be used as the basis for numerous non-Kolmogorovean models.

3. Negative probabilities are well defined on the mathematical level of rigorousness.

4. Pathological (or nonclassical) behaviour of 'quantum probabilities' (and, in particular, Bell's inequality) is a consequence of the formal use of

Kolmogorov's probability model.

5. Bell's inequality could not be used as an argument for *nonlocality* or *nonreality*. In may be that physical reality nonlocal or nonobjective. However, Bell's inequality has nothing to do with these problems.

6. The Wiegner distribution is well defined both in the ensemble and frequency frameworks.

7. From the frequency viewpoint non-Kolmogorovean probabilistic behaviour is (typically) the exhibition of the violation of the law of large numbers.

8. From the ensemble viewpoint non-Kolmogorovean probabilistic behaviour is a consequence of the use of ensembles of infinitely large 'volume'. There is no statistical *reproducibility* of properties for finite approximations of these infinite ensembles.

9. Quantum states (wave functions) describe such infinite (ideal) ensembles with statistical nonreproducibility of properties.

A large number of mathematicians and physicists took part in the discussion of results exposed in this book. I want to use the opportunity to express my deepest gratitude to all of them. I feel myself especially indebted to L. Accardi, S. Albeverio, H. Atmanspacher, Z. Hradil, W. de Muynck, H. Rauch, J. Summhammer for fruitful discussions.

Växjö–Clermont-Ferrand–Tokyo, 1998-99.

Chapter 1

Foundations of Probability Theory

There is no 'general probability theory'. There exist an incredible number of different mathematical probabilistic formalisms [74]–[78], [15], [16], [40], [23], [39], [50], [86]–[88], [1], [42], [92], [60], and, moreover, each of these formalisms has a few different interpretations. We shall discuss some of these theories which will be useful in further physical considerations.

1 A few words about measures

We recall some notions of measure theory. A system F of subsets of a set Ω is called an *algebra* if the sets $\emptyset, \Omega$ belong to F and the union, intersection and difference of two sets of F also belong to F. In particular, for any $A \in F$, a *complement* $\bar{A} = \Omega \setminus A$ of A belongs to F. Denote by F_Ω the family of all subsets of Ω. This is the simplest example of an algebra.

Let F be an algebra. A map $\mu : F \to \mathbf{R}_+$ is said to be a *measure* if $\mu(A \cup B) = \mu(A) + \mu(B)$ for $A, B \in F, A \cap B = \emptyset$. A measure μ is called σ-additive if, for every sequence $\{A_n\}_{n=1}^\infty$ of sets $A_n \in F$ such that their union $A = \cup_{n=1}^\infty A_n$ belongs to F, $\mu(A) = \sum_{n=1}^\infty \mu(A_n)$.

An algebra $\mathcal{F}$ is said to be a σ-algebra if, for every sequence $\{A_n\}_{n=1}^\infty$ of sets $A_n \in \mathcal{F}$, their union $A = \cup_{n=1}^\infty A_n$ belongs to $\mathcal{F}$.

Let Ω_1, Ω_2 be arbitrary sets and let G_1, G_2 be some systems of subsets of Ω_1 and Ω_2, respectively. A map $\xi : \Omega_1 \to \Omega_2$ is called *measurable* (or more precisely $((\Omega_1, G_1), (\Omega_2, G_2))$-measurable) if, for any set $A \in G_2$, the

set $\xi^{-1}(A) \in G_1$. We shall use the notation $\xi : (\Omega_1, G_1) \to (\Omega_2, G_2)$ to indicate the dependence on G_1, G_2. Typically we shall consider measurability of maps in the case in that $G_j, j = 1, 2$, are algebras or σ-algebras.

Let A be a set. A *characteristic function* I_A of the set A is defined as $I_A(x) = 1, x \in A$, and $I_A(x) = 0, x \in \bar{A}$.

Let $A = \{a_1, ..., a_n\}$ be a finite set. We shall denote the *cardinality* n of A by the symbol $|A|$.

2 Classical and ensemble definitions of probability

1. Classical definition of probability. The theory of probability originated from the study of problems connected with ordinary games of chance. In all these games the results that are *a priori* possible may be arranged in a finite number of cases assumed to be perfectly symmetrical, such as the cases represented by the six sides of a dice, the 52 cards in an ordinary pack of cards, and so on. This fact seemed to provide a basis for a rational explanation of the observed stability of statistical frequencies, and the 18th century mathematicians were thus led to the introduction of the famous *principle of equally possible cases.* According to this principle, a division into equally possible cases is possible in all random experiments, and the probability of an event is defined as the ratio between the number of cases favorable to the event, and the total number of possible cases. The main disadvantage of this probability theory is that the idea of symmetry cannot be applied to all random phenomena. For example, the classical definition of probability describes only a symmetric coin or dice. This definition cannot be used in the case of a violation of symmetry (see von Mises [88] for an extended critique of the classical definition). Denote by C the set of all possible cases. The classical theory operated on finite sets $C = \{c_1, ..., c_N\}$. For example, if a dice is considered, then $C = \{1, ..., 6\}$. Let E belong to the algebra F_C of all subsets of the set C. Then classical probability is defined by the equality

$$\mathbf{P}(E) = |E|/|C|. \tag{2.1}$$

The map $\mathbf{P} : F_C \to T_C \subset \mathbf{R}_+$, where $T_C = \{x = k/N : k = 0, 1, ..., N\}, N = |C|$, is a measure and $\mathbf{P}(C) = 1$. This measure is *uniform:* $\mathbf{P}(\{c_j\}) = 1/N$ and $\mathbf{P}(E) = \frac{1}{N} \sum_{c_j \in E} 1$.

We could not use (2.1) for infinite sets C in the framework of real analysis (there are no actual infinities in $\mathbf{R}$). This problem seems to be solved on the basis of the Kolmogorov measure-theoretic approach. But the classical definition (2.1) is not preserved in that approach. There are other possibilities to extend the classical definition of probability to infinite sets C. In principle we need not identify the set T_C of values of the classical probability with a subset of the set $\mathbf{R}$ of real numbers. It can be considered as just a subset of the set $\mathbf{Q}$ of rational numbers. It would be possible to extend the classical definition of probability by identifying T_C with a subset of other number system X such that $\mathbf{Q} \subset X$, see Chapter 4.

2. Ensemble (proportional) definition of probability. We start with the following classical example. There is an urn which contains balls of two colours, black and white. Let N_b and N_w be respectively the numbers of black and white balls; $N = N_b + N_w$ is the total number of balls in the urn. By definition a probability is the coefficient of the proportion between the number of balls of the concrete colour and the total number of balls: $\mathbf{P}(b) = \frac{N_b}{N}$ and $\mathbf{P}(w) = \frac{N_w}{N}$. In the general case we have a finite set S (an ensemble). Elements s of S have some properties. Denote the set of these properties by π_S. Each property $\xi \in \pi_S$ can be described as a map $\xi : S \to K_\xi$, where $K_\xi = \{1, 2, ..., k_\xi\}$ is a finite set (a numerical cod of the property ξ). We set $S(\xi = j) = \{s \in S : \xi(s) = j\}$; denote by $F(\pi_S)$ the collection of all these sets. By definition these are *events* and their probability is defined by

$$\mathbf{P}(S(\xi = j)) = \frac{|S(\xi = j)|}{|S|}. \tag{2.2}$$

If we assume that $F(\pi_S)$ is an algebra of sets then the map $\mathbf{P} : F(\pi_S) \to T_S \subset \mathbf{R}_+$, where $T_S = \{x = k/N : k = 0, 1, ..., N\}$ and $N = |S|$, is a measure and $\mathbf{P}(S) = 1$. If all one point sets s belong to the algebra $F(\pi_S)$, then $F(\pi_S)$ is the algebra of all subsets of S (i.e., $F(\pi_S) = F_S$) and $\mathbf{P}$ is the uniform distribution: $\mathbf{P}(\{s\}) = 1/N$. In this case we can connect the ensemble (proportional) definition with the classical definition: *the elements of the ensemble S can be interpreted as equally possible cases.*

The conditional probabilities will play an essential role in further quantum considerations. Now we demonstrate how these probabilities are introduced in the ensemble approach. Let $B = S(\xi = l), A = S(\eta = k), \xi, \eta \in \pi_S$. Let the set $C = A \cap B \in F(\pi_S)$. This means that there exists a property $\theta \in \pi_S$ such that $C = S(\theta = m)$. Conditional probability of the event B under the

condition A is defined as

$$\mathbf{P}_S(B/A) \equiv \mathbf{P}_A(B) = |B \cap A|/|A|$$

(we must extract from the ensemble S the sub-ensemble A and find the proportion of elements $s \in A$ which has the property $\xi(s) = l$). Thus we can easily obtain that

$$\mathbf{P}_S(B/A) = \mathbf{P}_S(B \cap A)/\mathbf{P}_S(A), \ \mathbf{P}_S(A) > 0. \tag{2.3}$$

This is well known Bayes' formula. We note that in the ensemble framework it is a theorem. In standard textbooks the ensemble index is omitted:

$$\mathbf{P}(B/A) = \mathbf{P}(B \cap A)/\mathbf{P}(A), \mathbf{P}(A) > 0. \tag{2.4}$$

Remark 2.1. If $F(\pi_S)$ is not an algebra, then $A, B \in F(\pi_S)$ need not imply that $C = A \cap B \in F(\pi_S)$. In this case we could not use Bayes' formula (2.4). Moreover, in such a case it is insensible to speak about conditional probabilities. There is no property θ of elements s of S such that $C = S(\theta = m)$. Thus the set $C = \{s \in S : \xi(s) = l\} \cap \{s \in S : \eta = k\}$ cannot be described by properties of S. From the physical viewpoint it means that we could not verify two properties ξ and η simultaneously. If we try to extract the sub-ensemble A from S by verifying the property η, then we change the property ξ of $s \in S$.

As a simple consequence of (2.4) we obtain another important formula:

$$\mathbf{P}(A \cap B) = \mathbf{P}(B/A)\mathbf{P}(A). \tag{2.5}$$

By symmetry we find

$$\mathbf{P}(A \cap B) = \mathbf{P}(A/B)\mathbf{P}(B). \tag{2.6}$$

Thus we have:

$$\mathbf{P}(A/B) = \frac{\mathbf{P}(B/A)\mathbf{P}(A)}{\mathbf{P}(B)}. \tag{2.7}$$

To be more careful, we have to indicate the dependence of probabilities on corresponding ensembles: $\mathbf{P}_B(A) = \frac{\mathbf{P}_A(B)\mathbf{P}_S(A)}{\mathbf{P}_S(B)}$.

In further quantum considerations we shall often use the following consequence of Bayes' formula. Let $A_k \in F(\pi_S), k = 1, ..., m, \cup_{k=1}^m A_k = S$ and $A_k \cap A_l = \emptyset, k \neq l$. Then, for every $C \in F(\pi_S)$ such that $C \cap A_k \in F(\pi_S)$, we have:

$$\mathbf{P}_S(C) = \sum_{k=1}^m \mathbf{P}_S(A_k)\mathbf{P}_{A_k}(C).$$

It is the well known formula of *total probability.* In standard textbooks this formula is written as

$$\mathbf{P}(C) = \sum_{k=1}^{m} \mathbf{P}(A_k)\mathbf{P}(C/A_k).$$

Thus concrete ensembles which are used to define left and right hand sides probabilities are not taken into account. We shall see that in quantum formalism this manipulation with the ensemble index will imply such unexpected consequences as *non-locality of space-time and super-luminal signals and death of reality.*

The direct generalization of proportional formula (2.2) for ensemble probabilities to infinite ensembles S is impossible in the framework of real analysis, because there are no actual infinities (infinitely large numbers) in the field of real numbers $\mathbf{R}$. A measure-theoretical approach (see section 4) provides some indirect generalization. However, this measure-theoretical approach is not the unique possibility to extend the proportional definition of probability to infinite ensembles. In Chapter 4 we shall consider ensembles which have structures of trees with an infinite number of vertexes (with p branches leaving each vertex; there $p > 1$ is a prime number). For such ensembles we can directly use (2.2) to define ensemble probabilities (there $N = |S|$ can be an infinite large number belonging to the field of so called p-adic numbers). Other possibility for extending (2.2) to infinite ensembles S is to use nonstandard analysis (see [3]).

3 Frequency theory of probability

This theory was the first where the principle of the stabilization of statistical frequencies was realized on a mathematical level. In fact, this principle was used as the definition of probability. Let us recall the main notions of a frequency theory of probability [86]–[88] of Richard von Mises (1919).[1] This theory is based on the notion of a collective. Consider a random experiment $\mathcal{S}$ and denote by $L = \{s_1, ..., s_m\}$ the set of all possible results of this experiment. The set S is said to be the label set, or the set of attributes. We consider only finite sets L. Let us consider N realizations of $\mathcal{S}$ and write a

[1] In fact, already in 1866 John Venn, see [105], tried to define a probability explicitly in terms of relative frequencies.

result x_j after each realization. Then we obtain the finite sample:

$$x = (x_1, ..., x_N),\ x_j \in L. \tag{3.1}$$

A *collective* is an infinite idealization of this finite sample :

$$x = (x_1, ..., x_N, ...),\ x_j \in L, \tag{3.2}$$

for which the following two von Mises' principles are valid.

The first is the *statistical stabilization of relative frequencies* of each attribute $\alpha \in S$ in the sequence (3.2). Let us compute frequencies $\nu_N(\alpha; x) = n_N(\alpha; x)/N$ where $n_N(\alpha; x)$ is the number of realizations of the attribute α in the first N tests. The principle of the statistical stabilization of relative frequencies says : **the frequency $\nu_N(\alpha; x)$ approaches a limit as N approaches infinity for every label** $\alpha \in L$. This limit $\mathbf{P}(\alpha) = \lim \nu_N(\alpha; x)$ is said to be the probability of the label α in the frequency theory of probability. Sometimes this probability will be denoted by $\mathbf{P}_x(\alpha)$ (to show a dependence on the collective x).

"We will say that a collective is a mass phenomenon or a repetitive event, or simply a long sequence of observations for which there are sufficient reasons to believe that the relative frequency of the observed attribute would tend to a fixed limit if the observations were infinitely continued. This limit will be called the probability of the attribute considered within the given collective" [87].

The second principle is the so-called principle of *randomness*. Heuristically it is evident that we cannot consider, for example, the sequence $z = (0, 1, 0, 1, ..., 0, 1, ...)$ as a random object (generated by a statistical experiment). However, the principle of the statistical stabilization holds for z and $\mathbf{P}(0) = \mathbf{P}(1) = 1/2$. Thus, we need an additional restriction for sequences (3.2). This condition was proposed by von Mises:

The limits of relative frequencies have to be stable with respect to a place selection (a choice of a subsequence) in (3.2).

In particular, z does not satisfy this principle. For example, if we choose only even places, then we obtain the zero sequence $z_0 = (0, 0, ...)$ where $\mathbf{P}(0) = 1, \mathbf{P}(1) = 0$.

However, this very natural notion was the hidden bomb in the foundations of von Mises' theory. The main problem was to define a class of place selections which induces a fruitful theory. The main and very natural restriction

is that a place selection in (3.2) cannot be based on the use of attributes of elements. For example, we cannot consider a subsequence of (3.2) constructed by choosing elements with the fixed label $\alpha_k \in L$. Von Mises proposed the following definition of a place selection:

(PS) "a subsequence has been derived by a place selection if the decision to retain or reject the nth element of the original sequence depends on the number n and on label values $x_1, ..., x_{n-1}$ of the $(n-1)$ presiding elements, and not on the label value of the nth element or any following element",

see [87], p.9. Thus a place selection can be defined by a set of functions $f_1, f_2(x_1), f_3(x_1, x_2), f_4(x_1, x_2, x_3), ...$, each function yielding the values 0 (rejecting the nth element) or 1 (retaining the nth element).

Here are some examples of place selections: (1) choose those x_n for which n is prime; (2) choose those x_n which follow the word 01; (3) toss a (different) coin; choose x_n if the nth toss yields heads. The first two selection procedures may be called *lawlike,* the third random. It is more or less obvious that all of these procedures are place selections: the value of x_n is not used in determining whether to choose x_n.

The principle of randomness ensures that no strategy using a place selection rule can select a subsequence that allows different odds for gambling than a sequence that is selected by flipping a fair coin. This principle can be called the *law of excluded gambling strategy.*

The definition (PS) induced some mathematical problems. If a class of place selections is too extended then the notion of the collective is too restricted (in fact, there are no sequences where probabilities are invariant with respect to all place selections). This was the main point of criticism of von Mises' theory. This problem has been investigated since the 1930s and solved only in the 1970s on the basis of Kolmogorov's notion of algorithmic complexity [76].

However, von Mises himself was satisfied by the following operational solution of this problem. He proposed [88] to fix for any collective a class of place selections which depends on the physical problem described by this collective. Thus he removed this problem outside the mathematical framework.

The frequency theory of probability is not, in fact, the calculus of probabilities, but it is the calculus of collectives which generates the corresponding calculus of probabilities. We briefly discuss some of the basic operations for collectives (see [88] for the details).

As probability is defined on the basis of the principle of the statistical stabilization of relative frequencies, it is possible to develop quite fruitful

probabilistic calculus by using only this principle. Sequence (3.2) which satisfies the principle of the statistical stabilization of relative frequencies is said to be a *S-sequence.* Thus limits of relative frequencies in a S-sequence x need not be invariant with respect to some class of place selections [2].

(a) *Mixing and additivity.* Let x be a collective with the (finite) label space L_x and let $E = \{\alpha_{i_1}, ..., \alpha_{i_l}\}$ be a subset of L_x. The sequence (3.2) of x is transformed into a new sequence y_E by the following rule. If $x_j \in E$ then we write 1; if $x_j \notin E$ then we write 0. Thus the label set $L_{y_E} = \{0, 1\}$. It is easy to show that this sequence has the property of statistical stabilization for its labels. For example,

$$\mathbf{P}_{y_E}(1) = \lim \nu_N(E; x) = \lim \sum_{k=1}^{l} \nu_N(\alpha_{i_k}; x) = \sum_{k=1}^{l} \mathbf{P}_x(\alpha_{i_k}), \tag{3.3}$$

where $\nu_N(E; x) \equiv \nu_N(1; y_E) = n_N(1; y_E)/N$ is the relative frequency of 1 in y_E. To obtain (3.3) we have only used the fact that the addition is a continuous operation on the field of real numbers $\mathbf{R}$. We can show that the sequence y_E also satisfies the principle of randomness, see [88]. Hence this is a new collective. By this operation any collective x generates a probability distribution on the algebra F_{L_x} of all subsets of $L_x : \mathbf{P}(E) = \mathbf{P}_{y_E}(1)$. Sometimes it will be convenient also to denote this probability distribution by $\mathbf{P}_x(E)$ to distinguish probabilities corresponding to different collectives. Now we find the properties of this probability. As $\mathbf{P}(E) = \lim \nu_N(E; x)$ and $0 \leq \nu_N(E) \leq 1$, then (by the elementary theorem of real analysis) $0 \leq \mathbf{P}(E) \leq 1$. Hence the probability must yield values in the segment $[0, 1]$. Further, as the collective y_{L_x} corresponding to the whole label set L_x does not contain zeros, we obtain that $\nu_N(L_x; x) \equiv \nu_N(1; y_{L_x}) \equiv 1$ and, consequently, $\mathbf{P}(L_x) = 1$. Finally by (3.3) we find that the set function $\mathbf{P} : F_{L_x} \to [0, 1]$ is additive. Thus $\mathbf{P}$ is a normalized measure on the algebra F_{L_x} which yields values in $[0, 1]$. We remark that all these considerations can be repeated for S-sequences.

(b) *Partition and conditional probabilities.* Let x be a collective and let $A \in F_{L_x}$ and $\mathbf{P}(A) \neq 0$. We derive a new sequence $z(A)$ by retaining only those elements of x which belong to A and discarding all other elements

[2] Of course, the use of S-sequences contradicts to the philosophy of the modern probability theory which is based on generalizations of Mises' principle of randomness (such as Kolmogorov complexity [76] and Martin-Löf [83] theory of statistical tests). However, it seems that all this machinery of randomness is not used in quantum physics. Experimentalists are only interested in the statistical stabilization of relative frequencies.

(thus the label set $L_{z(A)} = A$). This operation is obviously not a place selection, since the decision to retain or reject an element of x depends on the label of just this element. The sequence $z(A)$ is again a collective, see [88]. Suppose that $\alpha_j \in A$ and let y_A be the collective generated by x with the aid of the mixing operation. Then $\mathbf{P}_{z(A)}(\alpha_j) = \lim_{N\to\infty} \nu_N(\alpha_j; z(A)) = \lim_{k\to\infty} \nu_{N_k}(\alpha_j; z(A))$, where $N_k \to \infty$ is an arbitrary sequence. As $\mathbf{P}(A) \neq 0$ then $M_k = n_k(1; y_A) \to \infty$ (this is the number of labels belonging to A among the first k elements of x). Thus

$$\mathbf{P}_{z(A)}(\alpha_j) = \lim_{k\to\infty} \nu_{M_k}(\alpha_j; z(A)) = \lim_{k\to\infty} n_{M_k}(\alpha_j; z(A))/M_k$$

$$= \lim_{k\to\infty} [n_{M_k}(\alpha_j; z(A))/k] : [M_k/k] = \mathbf{P}_x(\alpha_j)/\mathbf{P}_x(A).$$

We have used the property that $n_{M_k}(\alpha_j; z(A))$, the number of α_j among first M_k elements of $z(A)$, is equal to $n_k(\alpha_j; x)$, the number of α_j among first k elements of x. The probability $\mathbf{P}_{z(A)}(\alpha_j)$ is the conditional probability of the label α_j if we know that a label belongs to A. It is denoted by $\mathbf{P}(\alpha_j/A) \equiv \mathbf{P}_x(\alpha_j/A)$. As a consequence of this formula we obtain Bayes' formula:

$$\mathbf{P}_{z(A)}(B) = \sum_{\alpha_j \in B\cap A} \mathbf{P}_x(\alpha_j/A)$$

$$= \sum_{\alpha_j \in B\cap A} \mathbf{P}_x(\alpha_j)/\mathbf{P}_x(A) = \mathbf{P}_x(B \cap A)/\mathbf{P}_x(A). \tag{3.4}$$

In fact, this formula connects probabilities defined with respect to different collectives. The left hand side probability is $\mathbf{P}_{z(A)}$ and the right hand side probabilities are $\mathbf{P}_x$. As in the case of the ensemble probability, sometimes we shall use the symbol $\mathbf{P}_A(B)$ instead of $\mathbf{P}(B/A)$. It useful to remark that $\mathbf{P}_A : F_{L_x} \to [0,1]$ is a measure normalized by 1. In particular, the probability $\mathbf{P}$ may be written as the conditional probability $\mathbf{P}_{L_x}$.

As in the ensemble framework, here we can also obtain the formula of total probability (2.8). Formula (2.8) is often applied in the wrong way: probabilities $\mathbf{P}(A_k)$ are found with respect to one collective and conditional probabilities $\mathbf{P}(C/A_k)$ with respect to other collective. To apply this formula in the right way we have to use the index of a collective:

$$\mathbf{P}_x(C) = \sum_{k=1}^{m} \mathbf{P}_x(A_k)\mathbf{P}_x(C/A_k). \tag{3.5}$$

Formulas (2.5)–(2.7) can be also easily obtained in the frequency framework.

Remark 3.1. The Bayes formula in the frequency framework is a consequence of the possibility of using the operation of partition for collectives. It should be noticed that from the physical point of view the operation of partition is a physical condition, which means that by extracting the collective $z(A)$ from the original collective x we do not change the property of belonging to B or not. If the physical system does not satisfy this condition, we cannot use the Bayes formula (3.4). This does not mean that we cannot define the conditional probability $\mathbf{P}_A(B)$. But we cannot use (3.4) to find this probability.

It is important to remark that the conditional probabilities in (2.7) are defined with respect to different collectives, $z(A)$ and $z(B)$. From the physical point of view the connection (2.7) between these probabilities is possible only for physical systems which satisfy conditions discussed in Remark 3.1.

It is evident that we can also consider countable sets of attributes $L_x = \{\alpha_1, \alpha_2, ..., \alpha_m, ...\}$. If we use the additional condition $\sum_{j=1}^{\infty} \mathbf{P}(\alpha_j) < \infty$ for the probabilities of labels then $\mathbf{P}$ is a (discrete) measure on F_{L_x}. Moreover, this measure is σ-additive. However, the generalization of the frequency theory of probability to 'continuous' sets of attributes is a nontrivial mathematical problem, see [88], [102].

4 Kolmogorov's measure-theoretical theory

The axiomatics of the modern probability theory was proposed by Andrei Nikolaevich Kolmogorov [74] in 1933 to provide a reasonable mathematical description of this theory. The basis of Kolmogorov axiomatics was prepared at the beginning of this century in France by investigations of Borel [15]–[16] and Frechet [40] on the measure-theoretic approach to probability. At the same time Kolmogorov used ideas of von Mises [86] about the frequency definition of probability (see remarks in [74]).

By the Kolmogorov axiomatics the *probability space* is defined as the triple $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$, where Ω is an arbitrary set (points ω of Ω are said to be *elementary events*), $\mathcal{F}$ is an arbitrary σ-algebra of subsets of Ω (elements of $\mathcal{F}$ are said to be *events*), $\mathbf{P}$ is a σ-additive measure on $\mathcal{F}$ which yields values in the segment $[0,1]$ of the real line and normalized by the condition $\mathbf{P}(\Omega) = 1$.

Random variables on $\mathcal{P}$ are defined as measurable functions $\xi : (\Omega, \mathcal{F}) \to$

$(\mathbf{R}, \mathcal{B})$, where $\mathcal{B}$ is the Borel σ-algebra on the real line[3]. We shall use the symbol $RV(\mathcal{P})$ to denote the space of random variables over $\mathcal{P}$. Probability distribution of $\xi \in RV(\mathcal{P})$ is defined as $\mathbf{P}_\xi(B) = \mathbf{P}(\xi^{-1}(B))$ for $B \in \mathcal{B}$. This is a σ-additive measure on the Borel σ-algebra.

A.N. Kolmogorov motivated additivity of probability by additivity of frequency probability (see formula (3.3)); he also used frequency reasons to take the segment $[0, 1]$ as the range of values of a probabilistic measure. On the other hand, the condition of σ-additivity was considered by Kolmogorov as an additional mathematical (technical) condition to provide a fruitful integration theory based on the Lebesgue integral. In fact, Kolmogorov started with finite additive probabilities defined on algebras of sets. The spaces with σ-additive probabilities defined on σ-algebras were called generalized probability space.

The Kolmogorov theory also contains the additional axiomatic definition of conditional probabilities. By definition $\mathbf{P}(B/A)$ is defined by formula (2.4). Kolmogorov did not give any motivation for this definition in his book [74]. However, as he gave a clear motivation of all other properties of $\mathbf{P}$ on the basis of the von Mises frequency theory, it seems to be that he used the same frequency reasons for (2.4). In Kolmogorov's model two events A and B are said to be independent if

$$\mathbf{P}(A \cap B) = \mathbf{P}(A)P(B) \tag{4.1}$$

or

$$\mathbf{P}(B/A) = P(B), \ \mathbf{P}(A) > 0. \tag{4.2}$$

In the standard framework of Lebesgue integration we start with a σ-additive measure μ defined on some algebra F and then μ is extended over the σ-algebra $\mathcal{F}$ generated by F (Borel σ-algebra). This extension procedure, which is well defined from the mathematical point of view, is not so innocent from the probabilistic point of view. Kolmogorov remarked: " Even if the sets (events) A of F can be interpreted as actual and (perhaps only approximately) observed events, it does not, of course, follow from this that the sets of $\mathcal{F}$ reasonably admit of such an interpretation. Thus there is the possibility that while a field of probability $(F, \mathbf{P})$ may be regarded as the image (idealized, however) of actual random events, the extended field of probability $(\mathcal{F}, \mathbf{P})$ will still remain merely a mathematical structure. Thus sets of $\mathcal{F}$ are merely ideal events to which

[3]Thus $\xi^{-1}(B) \in \mathcal{F}$ for every $B \in \mathcal{B}$.

nothing corresponds in the outside world. However, if reasoning which utilizes probabilities of such ideal events leads us to a determination of the probability of an actual event of F, then, from an empirical point of view also, this determination will automatically fail to be contradictory", see [74], p.17. It should be noticed that the adherents of Kolmogorov's measure-theoretical approach to probability theory did not pay large attention to these ideas of Kolmogorov. This implied that manipulations with abstract probabilities belonging to $\mathcal{F}$ were considered as real probabilistic investigations. Moreover, if we need not pay attention to the difference between real and abstract probabilities, we could in principle omit the concrete probabilistic model from our considerations and operate with 'events' belonging to abstract σ-algebras. This is the main problem of worldwide use of Kolmogorov's measure-theoretical approach.

Remark 4.1. For example, Cramer, who used the Kolmogorov axiomatics to create the mathematical theory of statistics, had another point of view on the problem of verification: "any probability assigned to a specific event must, in principle, be liable to verification" [23]. The question of verification was the cornerstone of the von Mises theory for the continuous label set S. He showed that in the case $L_x = \mathbf{R}$ (or $\mathbf{R}^n$) a probability measure of an event E has the frequency interpretation iff the measure of the boundary of E is equal to 0, [88].

On the other hand, Kolmogorov himself developed actively the viewpoint that probability theory is a purely mathematical theory. Therefore the concrete structure of set algebra (or σ-algebra) does not play any role in probabilistic considerations. In his manifest "General Measure Theory and Probability Calculus", 1929 (see [99]), he wrote: "To outline the context of theory, it suffices to single out from probability theory those elements that bring out its intrinsic logical structure, but have nothing to do with the specific meaning of theory."

Finally we remark that in Kolmogorov's approach Bayes' formula (2.4) is just the definition of a conditional probability. I like to underline this fact. I have the experience that many scientists working in applications of probability are sure that Bayes' formula is a theorem. But this is right only for ensemble and frequency approaches. On the other hand, the formula of total probability (2.8) is a theorem of the Kolmogorov's theory. Here it holds true for a countable family of sets $A_k \in \mathcal{F}, \mathbf{P}(A_k) > 0, k = 1, ...,$ such that $\cup_{k=1}^{\infty} A_k = \Omega$ and $A_k \cap A_l = \emptyset, k \neq l$: for every $C \in \mathcal{F}$, $\mathbf{P}(C) = \sum_{k=1}^{\infty} \mathbf{P}(A_k)\mathbf{P}(C/A_k)$. To obtain this formula, we need to use the σ-additivity of probability and the definition (Bayes' formula) of conditional probabilities.

5 Kolmogorov's ideas on probability

It should be noticed that before to create the system of axioms of probability theory, A. N. Kolmogorov discussed ([99], 1929) some examples of 'generalized probabilities' which could not be described by his axiomatics. Moreover, probably we need not call these objects 'generalized probabilities'. It seems more natural to call ordinary probabilities (described by Kolmogorov's axiomatics) 'restricted probabilities'.

There is other side of the common use of Kolmogorov's approach which is not so visible as the disappearance of concrete probabilistic spaces. This is the idea that only Lebesgue measurable sets could play some role in probabilistic considerations. Of course, this is a consequence of the fact that Kolmogorov discussed merely the Lebesgue extension [99] (or the Borel extension [74]). However, in principle some sets which are not Lebesgue measurable may appear in probabilistic models connected with some natural phenomena. We shall discuss such a model in Chapter 2. On the other hand, Kolmogorov discussed in [99] non-Lebesgue extension of the linear Lebesgue measure μ on the segment $[0,1]$, namely, the result of Banach that μ can be extended to a measure $\bar{\mu}$ defined on the (σ-)algebra $F_{[0,1]}$ all subsets of $[0,1]$. It seems to be that Kolmogorov considered this measure as a good candidate to be probability. He also considered multidimensional case and pointed out that an extension $\bar{\mu}$ on $F_{[0,1]^n}$ of the Lebesgue measure μ on $[0,1]^n$ can be obtained by using the metric equivalence of a cube $[0,1]^n, n = 2, 3, ...$, and the segment $[0,1]$. Then he mentioned that in the case $n > 2$ such a measure does not satisfy the principle of equality of the measure of congruent sets. This is a consequence of example on the decomposition of a sphere into three sets being congruent to the sum of two others to within a countable set (see, for example, [48] for the proof):

Theorem 5.1. *A sphere S can be decomposed into disjoint sets $S = A \cup B \cup C \cup Q$ such that: (i) the sets A, B, C are congruent to each other; (ii) the set $B \cup C$ is congruent to each of the sets A, B, C; (iii) Q is countable.*

We continue to study the question on a domain of definition of probability. As we have seen, the ensemble approach does not imply automatically that the system of sets (events) $F(\pi_S)$ (corresponding to properties π_S of the ensemble S) must be an algebra. On the other hand, if Kolmogorov's axiomatics is used, then we have to start with (at least) an algebra. However, there may be random phenomena which do not possess the structure of an algebra. Why the union $A \cup B$ of two events A, B must always be an event?

Why the complement $D = \Omega \setminus C$ of an event C must always be an event? It is interesting that, before to propose the general axiomatics of probability theory [74] (1933), Kolmogorov discussed the problem of a domain of definition of probability [99] (1929). At that time he had the viewpoint which coincided with our viewpoint: "It is also doubtful if a measure connected with some problem in probability calculus need be closed" (i.e., defined on an algebra). In [99] Kolmogorov pointed out that "one should not assume, however, that the existence of measures of two intersecting sets implies the existence of measure for their sum or difference: there are certain important measures without this property." In particular, he discussed the following example.

Example 5.1. (Density of natural numbers; see, for example, [45], [91]. for the details). For a subset $A \subset \mathbf{N}$ the quantity

$$\delta(A) = \lim_{N\to\infty} \frac{|A \cap \{1, ..., N\}|}{N},$$

is called the *density* of A if the limit exists. Let $\mathcal{G}_d$ denote the collection of all subsets of $\mathbf{N}$ which admit density. It is evident that each finite $A \subset \mathbf{N}$ belongs to $\mathcal{G}_d$ and $\delta(A) = 0$. It is also evident that each subset $B = \mathbf{N} \setminus A$, where A is finite, belongs to $\mathcal{G}_d$ and $\delta(B) = 1$ (in particular, $\mathbf{P}(\mathbf{N}) = 1$). The reader can easily find examples of sets $A \in \mathcal{G}_d$ such that $0 < \delta(A) < 1$.

Proposition 5.1. *Let $A_1, A_2 \in \mathcal{G}_d$ and $A_1 \cap A_2 = \emptyset$. Then $A_1 \cup A_2 \in \mathcal{G}_d$ and*

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2). \tag{5.1}$$

Proof. As $A_1 \cap A_2 = \emptyset$, then $|(A_1 \cup A_2) \cap \{1, ..., N\}| = |A_1 \cap \{1, ..., N\}| + |A_2 \cap \{1, ..., N\}|$. ∎

Proposition 5.2. *Let $A_1, A_2 \in \mathcal{G}_d$. The following conditions are equivalent:*

$$1) A_1 \cup A_2 \in \mathcal{G}_d; \quad 2) A_1 \cap A_2 \in \mathcal{G}_d;$$

$$3) A_1 \setminus A_2 \in \mathcal{G}_d; \quad 4) A_2 \setminus A_1 \in \mathcal{G}_d.$$

There are standard formulas:

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2); \tag{5.2}$$

$$\mathbf{P}(A_1 \setminus A_2) = \mathbf{P}(A_1) - \mathbf{P}(A_1 \cap A_2). \tag{5.3}$$

Proof. We have

$$|(A_1 \cup A_2) \cap \{1, ..., N\}| = |A_1 \cap \{1, ..., N\}| + |A_2 \cap \{1, ..., N\}| - |(A_1 \cap A_2) \cap \{1, ..., N\}|.$$

Therefore, if, for example, $A_1 \cap A_2 \in \mathcal{G}_d$ then there exists a limit of the right hand side. It implies $A_1 \cup A_2 \in \mathcal{G}_d$ and (5.2) holds. Other implications are proved in the same way. ■

It is possible to find sets $A, B \in \mathcal{G}_d$ such that, for example, $A \cap B \notin \mathcal{G}_d$. Let A be the set of even numbers. Take any subset $C \subset A$ which has no density. In fact, you can find C such that

$$\frac{1}{N}|C \cap \{1, 2, \cdots, N\}|$$

is oscillating. There happen two cases: $C \cap \{2n\} = \{2n\}$ or $= \emptyset$. Set

$$B = C \cup \{2n - 1 : C \cap \{2n\} = \emptyset\}$$

Then, both A and B have densities one half. But $A \cap B = C$ has no density. Thus $\mathcal{G}_d$ is not a set algebra.

In 1929 A.N. Kolmogorov wrote [99]: "It is not known whether every measure is closable. If closure is possible, then it is not necessarily closable in only one way. It would seem that it is very difficult to find a measure that closes the measure given by the density of natural numbers." We can prove that the density of natural numbers can be closed (extended on the algebra $F_{\mathbf{N}}$ of all subsets of $\mathbf{N}$), see Theorem 5.4.

To formalize our considerations on the density of natural numbers, we propose the following definition.

Definition 5.1. *A system of subsets $\mathcal{G}$ of a set Ω, which has the properties described by Proposition 5.2 and contains $\emptyset$ and Ω, is called a set semi-algebra.*

Definition 5.2. *A function $\mathbf{P} : \mathcal{G} \to [0, 1]$, where $\mathcal{G}$ is a semi-algebra, is said to be a probability semi-measure if it satisfy the additivity condition (5.1) and $\mathbf{P}(\Omega) = 1$.*

Definition 5.3. *The system $\mathcal{P} = (\Omega, \mathcal{G}, \mathbf{P})$, where $\mathbf{P}$ is a probability semi-measure on a semi-algebra $\mathcal{G}$, is called a semi-probability space.*

Unfortunately we could not say anything more about such a generalization of a probability space, because the theory of integration with respect to probability semi-measures is not well developed.

We present the simplest construction of an extension of a measure μ on the algebra of all subsets. This construction is based on a representation of μ by a continuous linear functional on some space of functions and the application of the Hahn-Banach theorem.

Theorem 5.2. *Let μ be a (finite additive) measure on an algebra F of subsets of Ω. Then there exists a finite-additive extension $\bar{\mu}$ of μ on the algebra F_Ω of all subsets of Ω.*

Proof. We introduce the space of bounded functions

$$B(\Omega) = \{f : \Omega \to \mathbf{R} : \ \|f\|_\infty = \sup_{\omega \in \Omega} |f(\omega)| < \infty\} .$$

This is a normed space. We set

$$\mathcal{L}(F) = \{f = \sum_{k=1}^{n} c_k I_{A_k}, \ c_k \in \mathbf{R}, \ A_k \in F, \cup_{k=1}^{n} A_k = \Omega\} ,$$

where I_A is the characteristic function of the set A. This is the normed subspace of the space $B(\Omega)$. We define a linear functional $l_\mu : \mathcal{L}(F) \to \mathbf{R}$ by

$$l_\mu(f)(\equiv \int_\Omega f(\omega)\mu(d\omega)) = \sum_{k=1}^{n} c_k \mu(A_k) \text{ for } f = \sum_{k=1}^{n} c_k I_{A_k}.$$

This functional is well defined (i.e., $l_\mu(f)$ does not depend on a representation of f). If

$$f = \sum_{k=1}^{n_1} c_k^1 I_{A_k^1} \text{ and } f = \sum_{k=1}^{n_2} c_k^2 I_{A_k^2} ,$$

then it is always possible to find sets $B_l \in F$ $(l = 1, ..., N)$, $B_l \cap B_m = \emptyset, l \neq m$, $\Omega = \cup_{l=1}^{N} B_l$, such that all sets $A_k^j, j = 1, 2$, are represented as unions of sets B_l (of course, here we use the **structure of an algebra**). We write

$$f = \sum_{l=1}^{N} d_l I_{B_l}, \quad \text{where} \ \ B_l \in F, \ B_l \cap B_m = \emptyset, l \neq m, \ \Omega = \cup_{l=1}^{N} B_l . \qquad (5.4)$$

By using finite-additivity of μ we obtain that

$$\sum_{k=1}^{n_1} c_k^1 \ \mu(A_k^1) = \sum_{l=1}^{N} d_l \ \mu(B_l) = \sum_{k=1}^{n_2} c_k^2 \ \mu(A_k^2) .$$

This functional is continuous (bounded) on the normed space $\mathcal{L}(F)$, because on the basis of representation (5.4) we obtain:

$$|l_\mu(f)| = |\sum_{l=1}^{N} d_l \ \mu(B_l)| \leq \max_l |d_l| \sum_{l=1}^{N} \mu(B_l) = \mu(\Omega) \sup_{\omega \in \Omega} |f(\omega)| = \mu(\Omega) \|f\|_\infty .$$

In fact, the norm $\|l_\mu\| = \sup\{|l_\mu(f)| : \|f\|_\infty \leq 1,\ f \in \mathcal{L}(F)\}$ of l_μ is equal to $\mu(\Omega)$.

We apply the following well known theorem of functional analysis.

Theorem 5.3. (Hahn-Banach) *Let E be a normed linear space and let U be its linear subspace. Every continuous linear functional $l : U \to \mathbf{R}$ can be extended to a continuous linear functional $L : E \to \mathbf{R}$ in such a way that norms of the functionals l and L coincide: $\|L_\mu\| = \|l_\mu\|$.*

An extension on the algebra F_Ω of the measure μ is defined by $\bar{\mu}(A) = L_\mu(A), A \in F_\Omega$, where $L_\mu : B(\Omega) \to \mathbf{R}$ is an extension of the continuous linear functional $l_\mu : \mathcal{L}(F) \to \mathbf{R}$ given by the Hahn-Banach theorem. Linearity of L_μ implies that the $\bar{\mu}$ is finite-additive.

Finally we have to prove that $\bar{\mu}(A) \geq 0$ for any $A \in F_\Omega$. Suppose that there exists $A \in F_\Omega$ such that $c = \bar{\mu}(A) < 0$. Then $\bar{\mu}(\bar{A}) = \mu(\Omega) - c > \mu(\Omega)$. On the other hand, $\bar{\mu}(\bar{A}) = L_\mu(I_{\bar{A}}) \leq \|I_{\bar{A}}\|_\infty \|L_\mu\| = \|l_\mu\| = \mu(\Omega)$. This contradiction implies that $\bar{\mu}$ is non-negative. ■

On the other hand, the $\bar{\mu}$ may be not σ-additive even if μ is σ-additive. It seems that an answer to the question:

"Is it possible in the general case to construct a σ-additive extension $\bar{\mu}$ on the algebra F_Ω of a σ-additive measure μ?"

is unknown.

Another difficulty is that the proof of the Hahn-Banach theorem is based on the *axiom of choice.* Therefore we also have to use this axiom to obtain an extension of probability. However, the place of the axiom of choice in quantum physics is not clear. Thus it is not easy to find the range of possible applications of probabilities extended on F_Ω with the aid of the Hahn-Banach theorem.

It seems that in general case it is impossible to obtain the existence of an extension $\bar{\mu}$ of μ without the axiom of choice.

However, the main problem is *non-uniqueness* of an extension $\bar{\mu}$. By our construction $\bar{\mu}$ is determined by an extension L_μ of the functional l_μ. In general such an extension is not unique.

Corollary 5.1. *Let $\mathbf{P}$ be a probabilistic measure on an algebra F of subsets of Ω. Then there exists a finite-additive extension $\bar{\mathbf{P}}$ of $\mathbf{P}$ on the algebra F_Ω of all subsets of Ω.*

Proof. By the Hahn-Banach theorem $1 = \mathbf{P}(\Omega) = \|l_\mu\| = \|L_\mu\|$. As, for each $A \in F_\Omega$, $\|I_A\|_\infty = 1$, we obtain that $\bar{\mathbf{P}}(A) = L_\mu(I_A) \leq \|L_\mu\| \|I_A\|_\infty \leq 1$. Thus $\bar{\mathbf{P}}$ a (finite-additive) probabilistic measure. ■

In some physical models we may use 'probabilities' defined on the algebra F_Ω of all subsets of Ω which are obtained via the Hahn-Banach theorem. As it has been noticed, in general these probabilities are not σ-additive. However, finite-additivity is merely a mathematical problem. The real problem is non-uniqueness of an extension $\bar{\mathbf{P}}$. For instance, we start with a σ-additive probability $\mathbf{P}$ defined on a σ-algebra $\mathcal{F}$. Let us assume that some events $A \in F_\Omega \setminus \mathcal{F}$ have a physical meaning. Let $\bar{\mathbf{P}}_1$ and $\bar{\mathbf{P}}_2$ be different extensions of $\mathbf{P}$ to the algebra F_Ω. In principle, $\bar{\mathbf{P}}_1(A) \neq \bar{\mathbf{P}}_2(A)$. As mathematical arguments are not sufficient to fix a 'probability', we need to use some additional physical arguments to obtain the 'right extension'.

It seems that the situation with nonuniqueness is even more complicated. As in the above considerations, let us start with a σ-additive probability $\mathbf{P}$ defined on a σ-algebra $\mathcal{F}$. Let $\mathbf{P}_{\mathrm{L}}$ be the Lebesgue extension of $\mathbf{P}$ on the σ-algebra $\mathcal{F}_L$ of Lebesgue measurable sets. The $\mathbf{P}_{\mathrm{L}}$ is the unique σ-additive extension of $\mathbf{P}$ on $\mathcal{F}_L$. On the other hand, there may exist finite-additive extensions $\bar{\mathbf{P}}$ of $\mathbf{P}$ on $\mathcal{F}_L$ which do not coincide with $\mathbf{P}_{\mathrm{L}}$. As we have discussed many times, the condition of σ-additivity is a purely mathematical condition. Therefore from the physical viewpoint there are no reasons to choose only the σ-additive extension. Thus the standard choice, $\mathbf{P}_{\mathrm{L}}(A)$, of probability for events $A \in \mathcal{F}_L$ does not seem so natural from the physical viewpoint. We think that some paradoxes in quantum formalism are a consequence of the common opinion that only the Lebesgue extension $\mathbf{P}_L$ gives 'right physical probability'. In particular, the proof of famous Bell's inequality is based on such an assumption. Thus the Eistein-Podolsky-Rosen paradox (see Chapter 2) might be a consequence of the conventional (but probably non-physical) choice of an extension of probability.

We have discussed norm-preserving extensions of probability obtained via the Hahn-Banach theorem. In principle there may exist extensions $L_\mu : B(\Omega) \to \mathbf{R}$ of the linear functional $l_\mu : \mathcal{L} \to \mathbf{R}$ which increase the norm: $\|L_\mu\| > \|l_\mu\|$. If we define an extension of a measure $\mu : F \to \mathbf{R}_+$ with the aid of such an extension, $\bar{\mu}(A) = L_\mu(I_A), A \in F_\Omega$, then we could not be sure that $\bar{\mu}$ is non-negative. In this way starting with probability $\mathbf{P} : F \to [0,1]$ we may obtain generalized probabilities $\bar{\mathbf{P}} : F_\Omega \to \mathbf{R}$ with negative values as well as with values which are larger than 1. We shall see in Chapter 3 that such generalized probabilities may have physical meaning. We note that if $\mathbf{P} : F \to [0,1]$ is a σ-additive probability, then it may be that a (norm-increasing) extension $\bar{\mathbf{P}} : F_\Omega \to \mathbf{R}$ is also σ-additive. In such a case we obtain a signed probability measure (a charge), see Chapter 3.

Moreover, it may exist a norm-increasing extension $\bar{\mathbf{P}}$ of σ-additive probability $\mathbf{P}$ (defined on a σ-algebra $\mathcal{F}$) on the Lebesgue σ-algebra $\mathcal{F}_L$. It may be that

$\bar{\mathbf{P}}(A) < 0$ (and $\bar{\mathbf{P}}(\bar{A}) > 1$) even for some $A \in \mathcal{F}_L$. Thus even for events $A \in \mathcal{F}_L$ which are typically considered as 'physical events', we could obtain negative generalized probabilities. Such negative probabilities have natural ensemble and frequency interpretations (see Chapter 3). There must be some special physical reasons to consider only the norm-preserving extension $\mathbf{P}_L$ of $\mathbf{P}$ on the Lebesgue σ-algebra $\mathcal{F}_L$. If there are no such reasons, then in principle we can use signed probabilities $\bar{\mathbf{P}} : \mathcal{F}_L \to \mathbf{R}$ for the description of a physical model. It seems (see Chapter 3) that there are physical reasons to use such signed probabilities (instead of the standard probability $\mathbf{P}_L$) in some physical models. In particular it may be that $\mathbf{P}_L(A) = \mathbf{P}_L(B) = \lambda$, but $\bar{\mathbf{P}}(A) \neq \bar{\mathbf{P}}(B)$. In such a case we can split the conventional probability λ into a set of generalized probabilities. In particular, zero conventional probability can be split into a set of generalized probabilities (see Chapters 3 and 4 for the details).

We turn back to the density of natural numbers.

Theorem 5.4. *The density of natural numbers $\delta : \mathcal{G}_d \to [0,1]$ can be extended to a finite-additive measure $\bar{\delta} : F_{\mathbf{N}} \to [0,1]$.*

Proof. We apply again the Hahn-Banach theorem. However, as $\mathcal{G}_d$ is not an algebra, we could not directly apply the scheme of the proof of Theorem 5.1. Denote by $LIM(\mathbf{N})$ the subspace of the normed space $B(\mathbf{N})$ (of all bounded functions $f : \mathbf{N} \to \mathbf{R}$) consisting of all functions f for which there exists the mean value:

$$l_\delta(f) = \lim_{n\to\infty} \frac{1}{n}(f(1) + ... + f(n)) \,.$$

The $l_\delta : LIM(\mathbf{N}) \to \mathbf{R}$ is a continuous linear functional and $l_\delta(I_A) = \delta(A)$ for each $A \in \mathcal{G}_d$. By the Hahn-Banach theorem l_δ can be extended to a continuous linear functional $L_\delta : B(\mathbf{N}) \to \mathbf{R}$ and $1 = \delta(\mathbf{N}) = \|l_\delta\| = \|L_\delta\|$. We set $\bar{\delta}(A) = L_\delta(I_A)$ for $A \in F_{\mathbf{N}}$. The linearity of L_δ implies that $\bar{\delta} : F_{\mathbf{N}} \to \mathbf{R}$ is additive. By the same reasons as in Theorem 5.2 we obtain that $\bar{\delta}$ is non-negative. ■

If $A \in F_{\mathbf{N}} \setminus \mathcal{G}_d$, then

$$\bar{\delta}(A) \neq \lim_{n\to\infty} \frac{|A \cap \{1, 2, ..., n\}|}{n} \,.$$

Thus the frequency verification of the event A is impossible (the principle of the statistical stabilization is violated; compare with Chapters 2 and 4).

An extension of δ from semi-algebra $\mathcal{G}_d$ on the σ-algebra $F_{\mathbf{N}}$ given by Theorem 5.3 is not unique. If in some physical model some sets $A \in F_{\mathbf{N}} \setminus \mathcal{G}_d$ are considered

as physical events, then there must be special physical reasons to choose one or another extension of δ. And we have to remember that the principle of the statistical stabilization is violated for events $A \in F_{\mathbf{N}} \setminus \mathcal{G}_d$. As in the case of measures defined on algebras, there may exist extensions L_δ of l_δ which do not preserve the norm: $\|L_\delta\| > 1$. In such a case an extension $\bar{\delta}$ corresponding to L_δ can take negative values.

In principle we might consider Theorem 5.4 as the answer to the question of A. N. Kolmogorov on a possibility to close the density of natural numbers δ. However, Kolmogorov wanted **to find** a measure which closes δ. Theorem 5.4 is not constructive (it is based on the axiom of choice). Really it does not give the answer to Kolmogorov's question. The construction which has been used in Theorem 5.4 could not be applied to an arbitrary semi-measure μ. Thus we do not know the answer to the question: "Is it possible to close an arbitrary semi-measure?" (compare with Kolmogorov, 1929).

6 Measure-theoretical approach and interpretations of probability

Now we are going to discuss possible probability interpretations of the Kolmogorov measure-theoretical approach (the mathematical theory of a special class of measures). As we have seen, the probabilistic measures can be associated with all probability models (classical, ensemble, and frequency). Therefore it is in principle possible to use the measure-theoretical formalism and classical, ensemble or frequency interpretation. However, A. N. Kolmogorov proposed not only a mathematical formalism but also an interpretation of this formalism. We shall start with this interpretation.

1. Ensemble-frequency interpretation. Kolmogorov interpreted a probability in the following way: "... we may assume that to an event A which may or may not occur under conditions Σ is assigned a real number $\mathbf{P}(A)$ which has the following characteristics: (a) one can be practically certain that if the complex of conditions Σ is repeated a large number of times, N, then if n be the number of occurrences of event A, the ratio n/N will differ very slightly from $\mathbf{P}(A)$; (b) if $\mathbf{P}(A)$ is very small, one can be practically certain that when conditions Σ are realized only once the event A would not occur at all". This interpretation is a mixture of the frequency and ensemble interpretations. In fact, (a) is the frequency interpretation and (b)

is the ensemble interpretation. However, we cannot identify Kolmogorov's interpretation with any of these interpretations (for example, we may not assume (see [88], p.5) that *each* infinite repetition of Σ will generate a collective). This mixture of interpretations generated some problems and played a negative role in applications of probability theory. Kolmogorov did not separate the proportion (measure) in ensemble and the frequency of realizations. Moreover, it seems to be that he often reduced the proportion in an ensemble to the proportion (2.1) for possible cases[4]. For example, he considered the experiment of tossing a coin twice and obtained a finite space of elementary events $\Omega = \{HH, HT, TH, TT\}$, where the labels H, T are used for the sides of a coin. I think that Kolmogorov understood very well the weakness of his interpretation. For this reason he considered this problem again 30 years later and proposed the theory of algorithmic complexity of random sequences [76]. However, the latter theory is nothing other than the attempt to justify the frequency probability theory of von Mises.

Remark 6.1. As the ensemble-frequency interpretation is based on both frequency and proportional arguments, the range of applications of Bayes' formula (2.4) is restricted by Remarks 2.1 and 3.1. In fact the Bayes formula is the additional postulate of the Kolmogorov axiomatics. In principle we can use the Kolmogorov theory (probability spaces) without Bayes' formula (2.4). This theory will describe the physical systems with a violation of (2.4). This framework was developed by Accardi [1]; we shall discuss it in the connection with quantum theory.

As we have seen, there are two (essentially different) contributions of Kolmogorov to probability theory. The first is the measure-theoretical approach and the second is the ensemble-frequency interpretation. The first is purely mathematical and the second is phemenological. Of course, it is possible to combine Kolmogorov's measure-theoretical formalism with other interpretations of probability. However, we have to pay attention to the problem that the use of a specific interpretation induces some restrictions to Kolmogorov's measure-theoretical approach. We start with the ensemble probability.

2. Measure-theoretical approach and the ensemble interpretation. Let S be an arbitrary (probably infinite) ensemble. Let $\mathcal{P}_S = (S, \mathcal{F}, \mathbf{P}_S)$ be Kolmogorov's probability space based on S. This space can be used for probabilistic analysis on S. However, we have to remember that the set π_S of properties can differ from the set of random variables $RV(\mathcal{P}_S)$. There

[4]There Kolmogorov followed the historical tradition.

can exist random variables $\xi \in RV(\mathcal{P}_S)$ (and, in particular, sets $A \in \mathcal{F}$) which are not properties of $s \in S$. As it was mentioned by Kolmogorov, analysis based on $\xi \in RV(\mathcal{P}_S)$ can give some results for elements $\eta \in \pi_S$ which have no real physical meaning. On the other hand, there can exist properties $\eta \in \pi_S$ which are not random variables (these are non-measurable maps η on $\mathcal{P}_S$). Other important thing is that all probability distributions depend on the ensemble S.

Let us consider the following example. Let $\mathcal{P}_j = (\Omega_j, \mathcal{F}_j, \mathbf{P}_j), j = 1, 2$, be Kolmogorov's probability spaces and let $\xi_j : \Omega_j \rightarrow \mathbf{R}$ be random variables with probability distributions $\mathbf{P}_{\xi_j}$. Then in Kolmogorov's formalism it is **always** possible to construct a probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ such that there are well defined random variables $\bar{\xi}_j \in RV(\mathcal{P}), j = 1, 2$, such that $\mathbf{P}_{\bar{\xi}_j} = \mathbf{P}_{\xi_j}$. We can simply set $\Omega = \Omega_1 \times \Omega_2$, $\mathcal{F} = \mathcal{F}_1 \otimes \mathcal{F}_2$, $\mathbf{P} = \mathbf{P}_1 \otimes \mathbf{P}_2$ and $\bar{\xi}_j(\omega_1, \omega_2) = \xi_j(\omega_j)$. However, it does not sound reasonable that we can do the same thing in the ensemble framework. Let $\Omega_1 = \Omega_2 = S$ and $\xi_j \in \pi_S, j = 1, 2$. In general it does not sensible to use the ensemble $\Omega = S \times S$ for representing properties of the original ensemble S.

3. Measure-theoretical approach and frequency probability. The original viewpoint of R. von Mises was that Kolmogorov's probability measure is nothing other than the probability distribution $\mathbf{P}_x$ (on the label set L_x) of a collective x. The Kolmogorov probability space in Mises' theory is chosen as $\mathcal{P}_x = (L_x, \mathcal{F}_{L_x}, \mathbf{P}_x)$ where in general case $\mathcal{F}_{L_x}$ is some σ-algebra of subsets of L_x. As we have already pointed out, in the continuous case not all sets $A \in \mathcal{F}_{L_x}$ have the frequency meaning. In particular, if the measure-theoretical approach is used for the description of the frequency phenomena, then the possibility of the frequency verification for events $A \in \mathcal{F}_{L_x}$ must be controlled. However, it is even more important to control continuously the dependence of a probability space on a collective.

Let us consider the following example (which is similar to the example considered in the ensemble framework). Let $x^j, j = 1, 2$, be two collectives with the label sets L_j and probability distributions $\mathbf{P}_{x^j}$. Let $\mathcal{P}_j = (L_j, \mathcal{F}_{L_j}, \mathbf{P}_{x^j}), j = 1, 2$, be the corresponding Kolmogorov's probability spaces. Let $A_j \in \mathcal{F}_{L_j}$. Suppose that somewhere we need to use conditional probability $\mathbf{P}(A_1/A_2)$. What is the meaning of Bayes' formula (2.4) in this case?

4. Measure-theoretical approach and ensemble-frequency interpretation. As we have already mentioned, typically Kolmogorov's measure-theoretical formalism on abstract probability spaces is used together with the

ensemble-frequency interpretation of probability. However, as in the cases of the ensemble and frequency theories, we must be careful with applications of the abstract measure-theoretical formalism. We study the question of a choice of a probability space for the concrete probability experiment.

The part (a) of the ensemble-frequency interpretation of probability implies that the space Ω must describe occurrences of events in very long sequences of repetitions of some condition Σ (in the mathematical formalism sequences can have infinite length). It seems that collectives can be used for the description of such a phenomenon. However, the part (b) is related to occurrences of events under a single realization of conditions Σ. Probability of a single realization is nonsense for collectives. Let us try to solve the contradiction between probability in a long sequence of repetitions of Σ and a single realization of Σ. We may consider the space C of all possible collectives which can be induced by repetitions of Σ. Then we may introduce on C a probability measure $\mathbf{P}$ (which seems to have the meaning of an ensemble probability for the ensemble C) that would provide a mathematical description of the part (b). The latter would mean that if $\mathbf{P}(A)$ is very small, then a single realization of A (in one of collectives $x \in C$) is practically impossible (from the ensemble viewpoint). However, in the standard formalism the space C of all collectives is not used as a space of elementary events Ω.[5] Instead of C, there is used the space $\Omega = L^\infty$ of all infinite sequences of labels $\alpha \in L$. Such a choice gives measure-theoretical advantages. However, this implies the consideration of sequences which have no probabilistic meaning.

We construct now Kolmogorov's probability measure $\mathbf{P}^{\mathrm{Kol}}$ on the space of sequences Ω which gives (as it is commonly accepted) the mathematical realization for (a) and (b). We start with the consideration of a symmetric coin (with sides denoted by symbols 0 and 1), $L = \{0, 1\}$. Here we can use the classical definition of probabilities as the starting point for the construction of $\mathbf{P}^{\mathrm{Kol}}$. As there are two equally possible cases, the classical probabilities $\mathbf{P}^{\mathrm{cl}}(0) = \mathbf{P}^{\mathrm{cl}}(1) = 1/2$. Now consider m trials of the coin and write all possible samples (3.2). At this point it seems that the formalism is developed in the same way as in the von Mises theory. However, the next step demonstrates the crucial difference between two the approaches. Denote by $S_m = L^m$ the set of all vectors of length m with coordinates $0, 1$. This set is considered as

[5]The constructive probability theory (see, for example, [79]) can be considered as an attempt to realize on the mathematical level of rigorousness the idea to use C as a space of elementary events.

a statistical ensemble. Thus, for $\mathbf{i} = (i_1, ..., i_m) \in S_m$, $\mathbf{P}^{\mathrm{ens}}(\mathbf{i}) \equiv \mathbf{P}_{S_m}(\mathbf{i}) = 1/|S_m| = 1/2^m$. Bernoulli proved the following mathematical result for these ensemble probabilities:

Theorem 6.1. (Bernoulli) *The larger m is, the larger is the proportion of those vectors in S_m in which the relative number of zeros (or of ones) deviates from 1/2 by less than a given ϵ.*

Obviously this is the result for proportional probability. But Bernoulli and most authors state this result as the result for the frequency probability: if one throws a 'true' coin long enough it is almost certain that the relative number of heads will deviate by less than ϵ from 1/2.

The Kolmogorov probability measure $\mathbf{P}^{\mathrm{Kol}}$ on the space of elementary events

$$\Omega = \{\omega = (\omega_1, ..., \omega_n, ...) : \omega_j \in L\},$$

where $L = \{0, 1\}$, will be defined with the aid of the ensemble probabilities $\mathbf{P}_{S_m}$. For $\mathbf{i} = (i_1, ..., i_m) \in S_m$, a *cylindrical subset* of Ω with the base $\mathbf{i}$ is defined as $B_{\mathbf{i}} = \{\omega \in \Omega : \omega_1 = i_1, ..., \omega_m = i_m\}$. We set $\mathbf{P}^{\mathrm{Kol}}(B_{\mathbf{i}}) = \mathbf{P}_{S_m}(\mathbf{i}) = 1/2^m$. Denote the σ-algebra generated by all cylindrical subsets by $\mathcal{F}$ (i.e., this is the minimal σ-algebra which contains all cylindrical subsets of Ω). $\mathbf{P}^{\mathrm{Kol}}$ is extended as a σ-additive measure on the σ-algebra $\mathcal{F}$.

It is typically assumed that the frequency part (a) of the interpretation can be described by following mathematical result for the measure $\mathbf{P}^{\mathrm{Kol}}$.

Theorem 6.2. (Law of large numbers) *For any $\epsilon > 0$,*

$$\mathbf{P}^{\mathrm{Kol}}(\{\omega \in \Omega : |\nu_m(1;\omega) - 1/2| > \epsilon\}) \to 0, m \to \infty,$$

where $\nu_m(1;\omega) = n_m(1;\omega)/m$ and $n_m(1;\omega) = \sum_{j=1}^{m} \omega_j$.

However, like the classical Bernoulli theorem, the law of large numbers is not connected with the frequency approximation of probabilities. This is the statement on the approximation of classical probabilities $\mathbf{P}^{\mathrm{cl}}(0) = \mathbf{P}^{cl}(1) = 1/2$ by ensemble probabilities. On the other hand, we could use the so called strong law of large numbers.

Theorem 6.3. (Strong law of large numbers) *There exists a subset $\Omega' \in \mathcal{F}$ such that $\mathbf{P}^{\mathrm{Kol}}(\Omega') = 1$ and $\nu_m(1,\omega) \to 1/2, m \to \infty$, for all sequences $\omega \in \Omega'$.*

But on the basis of this statement we could not say anything about the statistical stabilization of $\nu_m(1;\omega)$ for any concrete sequence $\omega \in \Omega$. The strong law of large numbers do not say anything about a frequency approximation of ensemble probabilities; this is the statement about the frequency

approximation of the classical probabilities $\mathbf{P}^{\mathrm{cl}}(0) = \mathbf{P}^{\mathrm{cl}}(1) = 1/2$ in the sense of the ensemble probabilities.

Conclusion. *The laws of large numbers cannot be applied for describing the statistical stabilization of frequencies in sampling experiments.*

We construct now Kolmogorov's measure $\mathbf{P}^{\mathrm{Kol}}$ in the general case. The classical definition of probability cannot be used for nonsymmetrical coin. We use the frequency definition. The statistical experiments for coin's tossing produce collectives x with the label sets $L = \{0, 1\}$. Let us assume that all these collectives have the same probability distribution: $q_0 = \mathbf{P}^{\mathrm{fr}}(0)$ and $q_1 = 1 - q_0 = \mathbf{P}^{\mathrm{fr}}(1)$. For cylindric set $B_{\mathbf{i}} = \{\omega \in \Omega : \omega_1 = i_1, ..., \omega_k = i_k\}, \mathbf{i} = (i_1, ..., i_k), i_l \in L = \{0, 1\}$, the probability is defined as

$$\mathbf{P}^{\mathrm{Kol}}(B_{\mathbf{i}}) = q_1^{|\mathbf{i}|} q_0^{k-|\mathbf{i}|}, \tag{6.1}$$

where $|\mathbf{i}| = i_1 + \cdots + i_k$.

In the symmetric case ($q_0 = q_1 = 1/2$) the origin of formula (6.1) has been explained in the ensemble framework. In the general case we could not apply the ensemble framework. Here we can apply frequency arguments. Let $x^{(j)} = (x_t^{(j)})_{t=1}^{\infty}, j = 1, 2, ..., k$, be collectives having the same label space $L = \{0, 1\}$ and probability distribution $\mathbf{P}_{x^{(j)}}(0) = q_0, \mathbf{P}_{x^{(j)}}(1) = q_1$. We form a new collective $x = (x_t)_{t=1}^{\infty}$ with the label space $L^k = L \times \cdots \times L$ by setting $x_t = (x_t^{(j)})_{j=1}^{k}, t = 1, 2, ...$ We assume that collectives $x^{(j)}$ are independent (see sections 9,10 for the details). In particular, this imply the factorization of the probability distribution $\mathbf{P}_x$ in a product of probability distributions $\mathbf{P}_{x^{(j)}}$. Thus, for each $\mathbf{i} = (i_1, ..., i_k), i_l = 0, 1$, there exists

$$\mathbf{P}_x(\mathbf{i}) = \lim_{M \to \infty} \nu_M(\mathbf{i}; x) = \prod_{l=1}^{k} \lim_{M \to \infty} \nu_M(i_l; x^{(l)}) = \prod_{l=1}^{k} \mathbf{P}_{x^{(l)}}(i_l), \tag{6.2}$$

where $\nu_M(\mathbf{i}; x) = n_M(\mathbf{i}; x)/M$ and $\nu_M(i_l; x^{(l)}) = n_M(i_l; x^{(l)})/M$ are relative frequencies for labels $\mathbf{i} \in L^k$ and $i_l \in L$ (in collectives x and $x^{(l)}$, respectively). Formula (6.2) can be used as the motivation for definition (6.1) of probability of a cylindric subset $B_{\mathbf{i}}$ of Ω.

The $\mathbf{P}^{\mathrm{Kol}}$ defined on cylindric subsets by (6.2) can be extended to a probability measure on the σ-algebra $\mathcal{F}$ of Ω generated by cylindric subsets.

We analyse now how the $\mathbf{P}^{\mathrm{Kol}}$ serves to purposes (a) and (b). Here we can also use the strong law of large numbers:

Theorem 6.4. (Strong law of large numbers for nonsymmetrical distributions) *There exists a subset* $\Omega' \in \mathcal{F}$ *such that* $\mathbf{P}^{\mathrm{Kol}}(\Omega') = 1$ *and* $\nu_M(\alpha;\omega) = n_M(\alpha;\omega)/M \to q_\alpha, M \to \infty, \alpha = 0, 1,$ *for all sequences* $\omega \in \Omega'$.

It seems that (with the same remarks as in the symmetric case) this is the mathematical realization of (a). The part (b) can be interpreted in the following way. If, for example, $q_0 << 1$ then, for each $j, \mathbf{P}^{\mathrm{Kol}}(\omega : \omega_j = 0) = q_0 << 1$. Thus 'probability to obtain 0 in the jth test is practically zero.' In fact, the problem is more complicated. In the nonsymmetrical case we could not interpret sequences $\omega \in \Omega$ (even some of them) as collectives generated by the statistical experiment[6]. The construction of Kolmogorov's measure $\mathbf{P}^{\mathrm{Kol}}$ demonstrates that $\omega \in \Omega$ have the meaning of (infinite) 'multi-labels' for the collective $x = (x_l)_{l=1}^\infty$, where $x_l = (x_l^{(j)})_{j=1}^\infty \in \Omega$, which is obtained on the basis of a sequence $\{x^{(j)}\}, j = 1, 2, ...,$ of independent collectives having the same probability distribution $q_\alpha, \alpha = 0, 1$ (i.e., a sequence (for $j = 1, 2, ...$) of parallel running sequences (for $l = 1, 2, ...$) of coins' tossings). Thus the strong law of large numbers says that 'practically all' these 'multi-labels' have the property of the statistical stabilization and limits of relative frequencies (accidentally!) coincide with probabilities q_α corresponding to collectives. Thus the probability measure $\mathbf{P}^{\mathrm{Kol}}$ describes the frequency approximation of probabilities only indirectly.

The $\mathbf{P}^{\mathrm{Kol}}$ describes only random phenomena which have the property of *ergodicity*. The ergodicity has the following meaning. First we consider the statistical experiment in that one person makes a long run of coin's tossings, $u = (u_1, ..., u_M, ...)$, and obtains the relative frequencies $\nu_M(\alpha; u) = n_M(\alpha; u)/M, \alpha = 0, 1$. Then we consider another statistical experiment in that all persons belonging to a large statistical ensemble S (population) make simultaneously just one coin's tossing. As the result of the latter experiment we obtain the proportions (in S), $\nu_S(\alpha) = |S(\alpha)|/|S|, \alpha = 0, 1,$ of persons who have obtained the label α. Then $\nu_M(\alpha; u) \approx \nu_S(\alpha)$ for large M and $|S|$. Of course, we could not assume that all random phenomena have the property of ergodicity. Thus in general the ensemble and frequency interpretations of probability must be separated.

Remark 6.2. Let collectives $x^{(j)}$ be independent, but not in general equally distributed: $\mathbf{P}_{x^{(j)}}(\alpha) = q_{\alpha j}, \alpha = 0, 1$. Then we obtain that $\mathbf{P}_x(\mathbf{i}) = \prod_{l=1}^N q_{i_l l}$. This

[6]Thus in the nonsymmetrical case the strong law of large numbers could not be interpreted in the same way as in the symmetric case. Only in the symmetric case we can interpret some of 'elementary events' $\omega \in \Omega$ as collectives generated by coin's tossing.

can be used as the motivation to define a probability of a cylindric subset of Ω by $\mathbf{P}^{\mathrm{Kol}}(B_{\mathrm{i}}) = \prod_{l=1}^{N} q_{i_l l}$. We underline again that there we could not use ensemble arguments to define probabilities of cylindric subsets.

Conclusion. *Kolmogorov's ensemble-frequency interpretation can be used only for ergodic random phenomena.*

7 Subjective (Bayesian) probability theory

According to the subjective interpretation of probability, it is the *degree of belief* in the occurrence of an event attributed by a given person at a given instant and with given set of information that is important. It is very important for our further quantum mechanical considerations that changing information changes probabilities. We illustrate this by an example.

Example 7.1. I have forgotten something : Have I sent a letter to my friend or not? I can propose my subjective probabilities q_1 (the letter was sent), q_2 (it was not sent), $q_1 + q_2 = 1, q_j \in [0, 1]$. Suppose that we have an ideal postal system, i.e., a letter could not disappear in the postal service. If I telephone to my friend and he tells me that he has received the letter, then at that moment the probabilities will immediately change: $q_1 \to 1$ and $q_2 \to 0$, in the opposite case: $q_1 \to 0$ and $q_2 \to 1$.[7]

In fact the subjective theory of probability is a sufficiently good theory from the operational point view. The main problem of this approach is how to choose the subjective probabilities in a concrete case. In this theory it is postulated that the probability depends on the *status of information* which is available to whoever evaluates probability. Thus the evaluation of probability is conditioned by some *a priori* ('theoretical') prejudices and by some facts ('experimental data'). However, in applications all this information is nothing other than information about frequency or proportional probabilities.

It must be noted that the subjective probability theory is described mathematically by the Kolmogorov probability space $(\Omega, \mathcal{F}, \mathbf{P})$. The Bayes formula (2.4) is the cornerstone of this theory (therefore, it is also called Bayesian theory). As we have discussed, in principle we can exclude (2.4) from the Kolmogorov theory and consider a more general formalism which describes violations of (2.4). Such an approach is impossible in the subjective framework.

[7]In quantum formalism such a reduction of subjective probabilities is nothing other than so called *collapse* of a wave function ($\phi = \sqrt{q_1}\phi_1 + \sqrt{q_2}\phi_2$).

The subjective probability theory is applied in the following form. There is a fixed set of hypotheses (events) $H_i \in \mathcal{F} : \cup_i H_i = \Omega, H_i \cap H_j = \emptyset,\ i \neq j$. Let $E \in \mathcal{F}$ be an event. Suppose that we know conditional probabilities $\mathbf{P}(E/H_i)$. Then we find $\mathbf{P}(H_i/E)$ by (3.8) and the formula of total probability: $\mathbf{P}(E) = \sum_i \mathbf{P}(E/H_i)\mathbf{P}(H_i)$, i.e.,

$$\mathbf{P}(H_i/E) = \frac{\mathbf{P}(E/H_i)\mathbf{P}(H_i)}{\sum_j \mathbf{P}(E/H_j)\mathbf{P}(H_j)}. \tag{7.1}$$

This is the standard form of *Bayes' theorem.*

Remark 7.1. Of course, Bayes' formula plays a great role in probability theory. However, as we have seen, there are restrictions for using this formula. These also are restrictions for using Bayesian probability theory. According to Bayesian theory $\mathbf{P}_H(E) = \mathbf{P}(E/H)$ is a subjective probability (a measure of an individual belief) on the basis of the known set of conditions H; in particular, $\mathbf{P}(E) = \mathbf{P}(E/\Omega)$ correspond to the set of all conditions. Therefore it is assumed that we can always extract the information H from the total amount of information Ω.

The main positive consequence of the subjective approach to probability theory is the connection between probability and information. The idea that probability is a measure of information on a random phenomenon (for example, a statistical ensemble) looks attractive. Typically such an information is reduced to our subjective knowledge about a random phenomenon. This information probability is coded by real number $q_{\mathrm{inf/pr}} \in [0, 1]$. Intuitively it is identified with the classical probability (based on the proportion of equally possible cases) or with the ensemble probability. However, the relation between subjective probability and classical or ensemble probabilities is indirect. The subjective probability approach claims that $q_{\mathrm{inf/pr}}$ is chosen on the basis of 'subjective reasons' of an individual.

The subjective probability approach can be strongly improved if we assume that 'subjective reasons' are nothing other than the calculation of probability with respect to an ensemble of ideas S (in the brain of an individual) which are connected with the concrete random phenomenon. Thus $q_{\mathrm{inf/pr}}(\alpha) = |S(\alpha)|/|S|$, where $S(\alpha)$ is the sub-ensemble of ideas which imply the property α. We shall study the connection between subjective (information) probability and probability on the space of ideas in Chapter 5 (in a p-adic information framework).

On the other hand, it seems natural to generalize subjective probability approach and construct information-probability theory in that (1) information which is coded in $q_{\mathrm{inf/pr}}$ is not considered as subjective information (i.e., $q_{\mathrm{inf/pr}}$ is an el-

ement of information reality which is not less objective that material physical reality); (2) $q_{\mathrm{inf/pr}}$ can be coded not only by real numbers belonging to $[0,1]$, but also by some other information vectors.

8 Foundations of Randomness

We study some special questions of the frequency probability theory connected with the principle of randomness (see, for example, [79], [103], [6] for the details).

1. Existence of collectives; Kamke's objection. As we have already remarked, the principle of randomness based on the invariance of limits of relative frequencies with respect to the set of *all possible place selections* is too general. In fact, there are no sequences which satisfy this principle. To show this, we follow arguments of E. Kamke [49] (see also [79]).

Let $L = \{0,1\}$ and $x = (x_j)_{j=1}^{\infty}$, $x_j \in L$, be a collective which induces the probability distribution $\mathbf{P}(\alpha) = 1/2$, $\alpha = 0, 1$. Consider the set SI of all strictly increasing sequences of natural numbers. This set can be formed independently of x; but, among elements of SI, we have the strictly increasing sequence $\{n : x_n = 1\}$. This sequence define a place selection which selects the subsequence $(11 \ldots 1 \ldots)$ from x. Hence x is not a collective after all!

The reader may well feel uncomfortable with the mathematical structure of the argument. Kamke claims to have shown that for every putative collection x there exists a place selection ϕ that disturbs the statistical stabilization of frequencies to probability 1/2. The use of the existential quantifier here classical (Platonistic). Indeed, it seems impossible to exhibit explicitly a procedure which satisfies von Mises' criterion (independence on value x_n) and at the same time selects the subsequence $(11 \ldots 1 \ldots)$ from x. The interesting analysis of this problem can be found in the review of M. van Lambalgen [103]. He is convinced that a satisfactory treatment of random sequences is possible only in set theories lacking the set power axiom, in which random sequences "are not already there." However, even we uncritical accept classical mathematics, Kamke's argument is somewhat beside the mark in that it fails to appreciate the purpose of von Mises' axiomatization. It refers to what **could** happen, whereas Mises' axioms are rooted in experience and refer to what **does** happen.

Remark 8.1. In various places von Mises likens the principle of randomness to the first law of thermodynamics. Both are statements of impossibility: the

principle of randomness is the principle of the excluded gambling strategy, while the first law (conservation of energy) is equivalent to the impossibility of a *perpetuum mobile* of the first kind. I think that such an analogy is not so natural. It would be more natural to connect the first law of thermodynamics with the first von Mises principle, the principle of the statistical stabilization of relative frequencies. The impossibility to perform precise measurements implies that the law of conservation of energy is only a statistical law. Thus it is just one of exhibitions of the principle of the statistical stabilization of relative frequencies. M. van Lambalgen compared the principle of randomness with the second law of thermodynamics, the law of increase of entropy or the impossibility of a perpetuum mobile of the second kind. Indeed, Kamke's objection is reminiscent of Maxwell's celebrating demon, that "very observant and neat-fingered being", invented to show that entropy decreasing evolutions may occur. Maxwell's argument of course in no way detracts from the validity of the second law, but serves to highlight the fact that statistical mechanics cannot provide an absolute foundation for entropy increase, since it does not talk about what **actually** happens (see [103] for further mathematical details).

The early attempts to formalize Mises' principle of randomness were based on considerations of different classes of lawlike place selection. The idea was to fix some class of lawlike place selections and then construct a set of collectives with respect to that class. Various authors (e.g. Popper, Reichenbach, Copeland) independently arrived at the so called *Bernoulli* selections. To discuss this class of place selections, it is convenient to formalize the definition of place selection.

Denote by L^* the set of all finite words $x = (x_1, \ldots, x_m)$, $x_j \in L$, $m = 1, 2, \ldots$ in the alphabet (label set) $L = \{\alpha_0, \alpha_1, ..., \alpha_l\}, l > 1$; as usual the symbol L^∞ is used to denote the set of all infinite sequences $x = (x_1, \ldots, x_m, \ldots)$, $x_j \in L$. Set $x_{1:n} = (x_1, \ldots, x_n)$ for $x \in L^\infty$ (this is the initial segment of the length n of the sequence x). A place selection ϕ is defined on the basis of a function $f : L^* \mapsto \{0, 1\}$. The domain of definition of a place selection ϕ corresponding to f is the set

$$\mathrm{dom}\, \phi = \{\mathrm{x} \in \mathrm{L}^\infty : \forall \mathrm{n}\ \exists \mathrm{k} \geq \mathrm{n} :\ \mathrm{f}(\mathrm{x}_{1:\mathrm{k}}) = 1\} \subset \mathrm{L}^\infty.$$

For $x \in \mathrm{dom}\ \phi$, we set $\phi(x) = \bigcap_n \bar{\phi}(x_{1:n})$, where the map $\bar{\phi} : L^* \mapsto L^*$ is defined as $\bar{\phi}(u\alpha) = \bar{\phi}(u)\alpha$ if $f(u) = 1$ and $\bar{\phi}(u\alpha) = \bar{\phi}(u)$ if $f(u) = 0$ (here $u = (u_1, \ldots, u_m)$ and $u\alpha = (u_1, \ldots, u_m, \alpha)$, $u_j, \alpha \in L$). Thus a place selection ϕ is a partial function $\phi : L^\infty \mapsto L^\infty$.

Example 8.1. (Bernoulli sequences) Let $w = (w_1, \ldots, w_s)$, $w_j \in L$, be a fixed word. For a sequence $x \in L^\infty$, we choose all x_n such that w is a

final segment of $x_{1:n}$. The domain of this place selection, ϕ_w, is the set of all sequences $x \in L^\infty$ which contain infinitely many occurrences of the word w. Formally ϕ_w is defined on the basis of the function $f_w : L^* \mapsto \{0,1\}$, $f_w(u) = 1$, if w is a final segment of u, $f_w(u) = 0$, if not. A *Bernoulli sequence* (with repect to a probability distribution $P(\alpha_j) = p_j$, $j = 1, \ldots, l$, $L = \{\alpha_1, \ldots, \alpha_l\}$) is a sequence $x \in L^\infty$ such that $\lim_{n\to\infty} \nu_n(\alpha_j; x) = p_j$, $j = 1, \ldots, l$, and for all words $w \in L^*$

$$\lim_{n\to\infty} \nu_n(\alpha_j; \phi_w x) = p_j, \; j = 1, \ldots, l, \tag{8.1}$$

where $\nu_n(\alpha_j; \phi_w x)$ is the relative frequency of occurrence of the label α_j in the initial segment of length n of the sequence $\phi_w x$. If $L = \{0,1\}$ is the binary alphabet and $\mathbf{P}_x(1) = p, \mathbf{P}_x(0) = 1 - p$, then (8.1) has the form

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j=1}^{n} (\phi_w x)_j = p$$

for all words $w \in \{0,1\}^*$.

The sets of Bernoulli place selections and sequences are denoted by symbols $\mathcal{U}_B$ and X_B, respectively.

A. Church [19] suggested to consider the set $\mathcal{U}_{\mathrm{Ch}}$ of place selections which are generated by total recursive functions $f : L^* \to \{0,1\}$ (functions which can be computed by using algorithms). Church's collectives (random sequences) are sequences $x \in L^\infty$ which satisfy the principle of the statistical stabilization and the principle of randomness for the set of place selections $\mathcal{U}_{\mathrm{Ch}}$. Denote the set of Church's collectives by the symbol X_{Ch}.

Both the sets $\mathcal{U}_B$ and $\mathcal{U}_{\mathrm{Ch}}$ are countable. The existence of Bernoulli sequences and Church's collectives is a consequence of the general result of A. Wald [110].

Let $p = (p_j) : \; p_j = \mathbf{P}(\alpha_j), j = 1, \ldots, l$, $L = \{\alpha_1, \ldots, \alpha_l\}$, be a probability distribution on the label set L. Let $\mathcal{U}$ be a set of place selections. We set

$$X(\mathcal{U}, p) = \{x \in L^\infty : \forall \phi \in \mathcal{U} \; \lim_{n\to\infty} \nu_n(\alpha_j; \phi x) = p_j, \; j = 1, \ldots, l\}.$$

Theorem 8.1. (Wald) *For any countable set $\mathcal{U}$ of place selections and any probability distribution p on the label set L, the set of sequences $X(\mathcal{U}, p)$ has the cardinality of the continuum.*

Thus at least for countable sets of place selections $\mathcal{U}$ Mises' frequency theory of probability can be developed on the mathematical level of rigorousness. R. von Mises was completely satisfied by this situation (see [88]). However, he was strongly against the idea to fix once and for all a set of place selections. By Mises the concrete set of place selections is determined by a physical problem. But mathematicians prefer to consider fixed classes of place selections. In particular, the large part of mathematical community consider Church's choice as the most reasonable. The author does not think that the choice of total recursive functions as place selections can be justified by some physical arguments. The idea that reality which can be studied by human mind can be reduced to reality produced by Turing machines looks rather primitive in the light of modern investigations of the processes of thinking. It seems that the brain uses transformations $\bar{\phi} : L^* \to L^*$ which based on non-recursive functions, [65], [66].

2. Geometric and frequency spaces. According to the modern ideology of geometry, geometric model is a pair (X, G), where X is a set of points and G is a group of transformations of X. Such an approach is closely connected with von Mises' approach to probability theory. Here we have a system of place selections $\mathcal{U}$ (which plays the role of a group of transformations G) and the space $X(\mathcal{U}, p)$ of 'probabilistic points'. The pair $(X(\mathcal{U}, p), \mathcal{U})$ can be called a *frequency probability model.* Moreover, as in geometry, we have to consider some algebraic structure on the system of transformations $\mathcal{U}$. We shall demonstrate that we have to use semigroups (with unit) of transformations $\mathcal{U}$.

Let $\mathcal{U}$ be a system of place selections containing the identity transformation. If $x \in X(\mathcal{U}, p)$, it is natural to assume that, for each $\phi \in \mathcal{U}$, $y = \phi x \in X(\mathcal{U}, p)$: each element ϕ of $\mathcal{U}$ transforms an $\mathcal{U}$-collective x in a new $\mathcal{U}$-collective. Thus, for each $\psi \in \mathcal{U}$, the sequence $z = \psi y = \psi \circ \phi x$ satisfies the principle of the statistical stabilization. Let $f = \psi \circ \phi \notin \mathcal{U}$. Then we can extend the system of transformations $\mathcal{U}$ by setting $\mathcal{U}' = \mathcal{U} \cup \{f\}$. It is evident that (under our assumption) the set of points $X(\mathcal{U}', p)$ coincides with the set $X(\mathcal{U}, p)$. Therefore it would be natural to assume from the beginning that $\mathcal{U}$ is a semigroup.

One of nice examples of frequency probability spaces is the space $(X_{\mathrm{Ch}}, \mathcal{U}_{\mathrm{Ch}})$ based on the system of totally recursive functions.

3. Ville's objection. Although Wald's reformulation of von Mises' ideas solved the problem of consistency, it lead to an objection of entirely different kind.

Theorem 8.2. (Ville, [106]) *Let $L = \{0,1\}$ and let $\mathcal{U} = \{\phi_n\}_{n=1}^{\infty}$ be a countable set of place selections. Then there exists $x \in L^\infty$ such that*

$$(a) \text{ for all } n, \lim_{N\to\infty} \frac{1}{N}\sum_{j=1}^{N}(\phi_n x)_j = \frac{1}{2};$$

$$(b) \text{ for all } N, \frac{1}{N}\sum_{j=1}^{N}(\phi_n x)_j \geq \frac{1}{2}.$$

Such an x is a collective with respect to $\mathcal{U}$ ($x \in X(\mathcal{U}, 1/2)$), but seems to be far too regular to be called random. Formally, x's with property (b) form a set of Lebesgue measure 0 (this is a consequence of the law of iterated logarithm).

4. Ensemble probability approach to randomness. Ville and Frechet used Theorem 8.2 to argue that collectives in the sense of von Mises and Wald do not necessarily satisfy all intuitively required properties of randomness. Ville introduced a new way of characterizing random sequences, based on the following idea: *a random sequence should satisfy all properties of probability one.* Strictly speaking, this is of course impossible: we have to choose countably many from among those properties. It must be underlined that Ville's idea is really completely foreign to von Mises. For von Mises, a collective $x \in L^\infty$ induces a probability on the set of labels L, not on the set of all sequences L^∞. Hence there is no connection at all between properties of probability one in L^∞ and properties of individual collectives.

Per Martin-Löf [83]–[85] proposed to consider 'recursive properties of probability one' (i.e., properties which can be tested with the aid of algorithms). Such an approach induces the fruitful theory of recursive tests for randomness (see, for example, [79], [112]). Similar approach was developed by Schnorr [98]. We underline that approaches of Martin-Löf and Schnorr (as well as Ville and Frechet) have nothing to do with the justification of Mises' frequency probability theory.

5. Kolmogorov Complexity. A.N. Kolmogorov tried to find foundations of randomness by reducing this notion to the notion of complexity. Let $L = \{0,1\}$ and $x \in L^*$.

Definition 8.1. (Kolmogorov) *Let $\mathcal{A}$ be an arbitrary algorithm. The complexity of a word x with respect to $\mathcal{A}$ is $K_{\mathcal{A}}(x) = \min l(\pi)$, where $\{\pi\}$ are the programs which are able to realize the word x with the aid of $\mathcal{A}$.*

Here $l(\pi)$ denotes the length of a program π. This definition depends on the structure of an algorithm $\mathcal{A}$. Later Kolmogorov proved the following theorem:

Theorem 8.3. *There exists an algorithm $\mathcal{A}_0$ (optimal algorithm) such that*

$$K_{\mathcal{A}_0}(x) \preceq K_{\mathcal{A}}(x) \tag{8.2}$$

for every algorithm $\mathcal{A}$.

As usual, (8.2) means that there exists a constant C such that $K_{\mathcal{A}_0}(x) \leq K_{\mathcal{A}}(x) + C$ for all words x. An optimal algorithm $\mathcal{A}_0$ is not unique.

Definition 8.2. *The complexity $K(x)$ of the word x is equal to the complexity $K_{\mathcal{A}_0}$ with respect to one fixed (for all considerations) optimal algorithm $\mathcal{A}_0$.*

The original idea of Kolmogorov [76], [77] was that complexity $K(x_{1:n})$ of initial segments $x_{1:n}$ of a random sequence x has to have the asymptotic $K(x_{1:n}) \sim n, n \to \infty$, i.e., we might not find a short code for $x_{1:n}$. However, this nice idea was rejected due to an objection of Per Martin Löf [85]. To discuss this objection and connection of Kolmogorov complexity with Martin Löf randomness, it is better to use conditional Kolmogorov complexity $K(x;n)$ instead of complexity $K(x)$. Complexity $K_{\mathcal{A}}(x;n)$ is defined as the length of a minimal program π which produces the output x on the basis of information that the length of the output x is equal to n.

Theorem 8.4. *Let f be a total recursive function such that $\sum_{n=1}^{\infty} 2^{-f(n)} = \infty$. Then, for every sequence x, $K(x_{1:n};n) < n - f(n)$ for infinitely many n.*

In particular, we can choose $f(n) = \log_2 n$. Thus, for any binary sequence x, $K(x_{1:n};n) < n - \log_2 n$ for infinitely many n. Hence 'Kolmogorov random sequences' do not exit.

P. Martin-Löf obtained also an estimate of $K(x_{1:n};n)$ from below:

Theorem 8.5. *Let f be such that $\sum_{n=1}^{\infty} 2^{-f(n)} < \infty$. Then, with probability one, $K(x_{1:n};n) \geq n - f(n)$ for all but finitely many n.*

In particular, we can choose $f(n) = 2\log_2 n$. Thus, for almost all binary sequences x, $K(x_{1:n};n) \geq n - 2\log_2 n$ for all but finitely many n. Therefore for almost all binary sequences Kolmogorov complexity $\phi(n) = K(x_{1:n};n)$ oscillates between graphs of the functions $g_{\max}(n) = n$ and $g_{\min}(n) = n - 2\log_2 n$ (with finitely many intersections with $g_{\min}(n)$). The graph of the function $\phi(n)$ has infinitely many intersections with the graph of the function $f_{\mathrm{mid}}(n) = \log_2 n$.

The following two theorems [85] give the connection between high Kol-

mogorov complexity (for infinitely many initial segments) and Martin-Löf randomness:

Theorem 8.6. *Let f be a total recursive function such that $\sum_{n=1}^{\infty} 2^{-f(n)}$ is recursively convergent. Then, if x is random in the sense of Martin-Löf, then $K(x_{1:n}; n) \geq n - f(n)$ for all but finitely many n.*

That $\sum_{n=1}^{\infty} 2^{-f(n)}$ is recursively convergent means that there is a recursive sequence $n_1, n_2, ..., n_k, ...$ such that

$$\sum_{n_m+1}^{\infty} 2^{-f(n)} \leq 2^{-m}, m = 1, 2, ...$$

Theorem 8.7. *If there exists a constant c such that $K(x_{1:n}; n) \geq n-c$ for infinitely many n, then the sequence x is random in the sense of Martin-Löf.*

Critical Remarks:

1) Despite of the great success of Kolmogorov and Martin-Löf approaches, it is doubtful that these approaches provide the adequate description of randomness in physical reality. The main objection is against the use of recursive functions (algorithms). On one hand, there are no reasons to suppose that random sequences produced by physical phenomena must pass all recursive tests for randomness (even the law of large numbers). On the other hand, 'randomness' of such sequences may be characterized by some systems of non-recursive transformations.

2) It seems impossible to reduce Martin-Löf randomness to Mises' randomness. Denote the class of Martin-Löf random sequences (with respect to the uniform distribution) by the symbol RM. The reduction of Martin-Löf randomness to Mises' randomness must be given by the equality $RM = X(\mathcal{U}, 1/2)$ for some class $\mathcal{U}$ of place selections. However, it seems impossible to find such a class $\mathcal{U}$. For example, let $\mathcal{U} = \mathcal{U}_{\text{Ch}}$ be the class (semigroup) of Church place selections. Then, as each $\phi \in \mathcal{U}_{\text{Ch}}$ gives a recursive property of probability one, we have $RM \subset X(\mathcal{U}, 1/2)$. Ville's result, combined with the observation that the Martin-Löf random sequences satisfy the law of the iterated logarithm, shows that the inclusion is strict. Moreover, it can be shown (see M. van Lambalgen [103]) that the set of sequences $X(\mathcal{U}, 1/2) \setminus RM$ is rather large.

Therefore approaches of Martin-Löf–Ville–Frechet and von Mises give totally different viewpoints to the notion of randomness. The first approach is based on the ensemble interpretation of probability and the second approach is based on the frequency interpretation of probability. As we have

already noticed, these interpretations could not be unified in one (mixed) ensemble-frequency interpretation.

9 Operation of combining of collectives

In the three basic operations discussed in section 3, one single collective x served each time as point of departure for the construction of a new collective. We consider the problem of combining of two or more given collectives. We start with S-sequences (sequences which satisfy the principle of the statistical stabilization).

Let $x = (x_j)$ and $y = (y_j)$ be two S-sequences with label sets L_x and L_y, respectively. We define a new sequence

$$z = (z_j), \; z_j = (x_j, y_j). \tag{9.1}$$

(in general z is not an S-sequence with respect to the label set $L_z = L_x \times L_y$). Let $a \in L_x$ and $b \in L_y$. Among the first N elements of z there are $n_N(a;z)$ elements with the first component equal to a. As $n_N(a;z) = n_N(a;x)$ is a number of $x_j = a$ among the first N elements of x, we obtain that $\lim_{N\to\infty} \frac{n_N(a;z)}{N} = \mathbf{P}_x(a)$. Among these $n_N(a;z)$ elements, there are a number, say $n_N(b/a;z)$ whose second component is equal to b. The frequency $\nu_N(a,b;z)$ of elements of the sequence z labeled (a,b) will then be

$$\frac{n_N(b/a;z)}{N} = \frac{n_N(b/a;z)}{n_N(a;z)} \frac{n_N(a;z)}{N}.$$

We set $\nu_N(b/a;z) = \frac{n_N(b/a;z)}{n_N(a;z)}$. Let us assume that, for each $a \in L_x$, the subsequence $y(a)$ of y which is obtained by choosing y_j such that $x_j = a$ is an S-sequence[8]. Then, for each $a \in L_x$, $b \in L_y$, there exists

$$\mathbf{P}_z(b/a) = \lim_{N\to\infty} \nu_N(b/a;z) = \lim_{N\to\infty} \nu_N(b;y(a)) = \mathbf{P}_{y(a)}(b).$$

We have

$$\sum_{b\in L_2} \mathbf{P}_z(b/a) = 1. \tag{9.2}$$

The existence of $\mathbf{P}_z(b/a)$ implies the existence of $\mathbf{P}_z(a,b) = \lim_{N\to\infty} \nu_N(a,b;z)$. Moreover, we have

$$\mathbf{P}_z(a,b) = \mathbf{P}_x(a)\mathbf{P}_z(b/a) \tag{9.3}$$

[8]In general such a choice of the subsequence $y(a)$ of y is not a place selection.

and $\mathbf{P}_z(b/a) = \mathbf{P}_z(a,b)/\mathbf{P}_x(a)$, if $\mathbf{P}_x(a) \neq 0$. By (9.2) and (9.3) we obtain

$$\sum_{a \in L_a} \sum_{b \in L_2} \mathbf{P}_z(a,b) = 1.$$

Thus in this case the sequence z is an S-sequence with the probability distribution $\mathbf{P}_z(a,b)$ defined by (9.3). The S-sequence y is said to be *combinable* with the S-sequence X. This relation is denoted by $\overleftarrow{xy}$. The relation of combining is a symmetric relation on the set of pears of S-sequences with strictly positive probability distributions. To show this, we write

$$\nu_N(a/b; z) = \frac{\nu_N(a,b;z)}{\nu_N(b;z)},$$

$a \in L_x$, $b \in L_y$. If $\overleftarrow{xy}$ and $\mathbf{P}_x(a) > 0$, $\mathbf{P}_y(b) > 0$, $a \in L_x$, $b \in L_y$, then, for each $b \in L_y$, $a \in L_x$, there exists

$$\mathbf{P}_z(a/b) = \lim_{N \to \infty} \nu_N(a/b; z) = \frac{\mathbf{P}_z(a,b)}{\mathbf{P}_y(b)} = \frac{\mathbf{P}_z(b/a)\mathbf{P}_x(a)}{\mathbf{P}_y(b)}.$$

Thus we obtain that $\overleftarrow{yx}$. On the other hand, if, for example, $\mathbf{P}_y(b) = 0$ and $\mathbf{P}_z(a,b) = 0$, then in principle $\nu_N(a/b; z)$ *may fluctuate.* In that case x is not combinable with y. The previous considerations can be summarized as the following proposition.

Proposition 9.1. *Let x and y be two S-sequences with strictly positive probability distributions. Then the following conditions are equivalent: 1) $\overleftarrow{xy}$; 2) $\overleftarrow{yx}$; 3) the sequence z defined by (9.1) is an S-sequence.*

If $\overleftarrow{yx}$ and $\overrightarrow{xy}$, then x and y are said to be combinable. This relation is denoted by $\overline{xy}$. If $\overleftarrow{xy}$, then the conditional probabilities $\mathbf{P}_z(b/a)$ are well defined even if $\mathbf{P}_x(a) = 0$. Typically such probabilities do not play any role in probabilistic considerations. We say that y is (mod $\mathbf{P}_{\mathrm{x}}$)–combinable with x, $\overleftarrow{xy}$ (mod $\mathbf{P}_{\mathrm{x}}$), if, for each $a \in L_x$, $\mathbf{P}_x(a) > 0$, the sequence $y(a)$ is an S-sequence. By Proposition 9.1 we have $\overleftarrow{yx}$ (mod $\mathbf{P}_{\mathrm{x}}$) $\Leftrightarrow$ z is an S-sequence $\Leftrightarrow$ $\overleftarrow{xy}$ (mod $\mathbf{P}_{\mathrm{y}}$). Thus we need not use the arrow to denote this relation of combining. This relation will be denoted as $\overline{xy}$ (mod $\mathbf{P}_{\mathrm{z}}$).

We introduce the operation of combining for collectives. We start with some preliminary considerations on place selections. Let $x = (x_j)$, $x_j \in L_x$, and $y = (y_j)$, $y_j \in L_y$, be two arbitrary sequences and let $z = (z_j)$, $z_j = (x_j, y_j)$. Let Φ_1, Φ_2 and G be some systems of place selections operated in

x, y and z, respectively. Let ϕ belongs to G : $\phi z = (z_{n_1}, z_{n_2}, ..., z_{n_k}, ...)$. We set $\phi^{(1)}x = (x_{n_j})$ and $\phi^{(2)}y = (y_{n_j})$. It should be noticed that in general $\phi^{(1)}$ and $\phi^{(1)}$ are not place selections in x and y, respectively[9]. We set $G_1 = \{f = \phi^{(1)} : \phi \in G\}$, $G_2 = \{g = \phi^{(2)} : \phi \in G\}$. Let $x = (x_j)$ and $y = (y_j)$ be Φ_1 and Φ_2 collectives, respectively. Let G be a system of place selections operated in $z = (z_j)$, $z_j = (x_j, y_j)$, such that, $\Phi_1 \subset G_1$ and $\Phi_2 \subset G_2$. If x and y are mod **P**-combinable as S-sequences, then they are said to be (mod $\mathbf{P}$, G)-combinable collectives if: (1) the limits $\mathbf{P}_x(a)$, $a \in L_x$, and $\mathbf{P}_y(b)$, $b \in L_y$, are insensitive to transformations belonging to G_1 and G_2, respectively; (2) the limits $\mathbf{P}_z(b/a)$, $\mathbf{P}_x(a) > 0$, and $\mathbf{P}_z(a/b)$, $\mathbf{P}_y(b) > 0$, are insensitive to place selections belonging to G. We can easily prove that x and y are (mod $\mathbf{P}$, G)-combinable collectives iff z is the G-collective.

Proof. 1) Let x and y be (mod $\mathbf{P}$, G)-combinable. Let $\phi \in G$. For $\mathbf{P}_x(a) > 0$, we have:

$$\mathbf{P}_{\phi z}(a, b) = \lim_{N\to\infty} \nu_N(a, b; \phi z) = \lim_{N\to\infty} \nu_N(b/a; \phi z)\nu_N(a; \phi z)$$

$$= \lim_{N\to\infty} \nu_N(b/a; \phi z)\nu_N(a; \phi^{(1)}x) = \mathbf{P}_z(b/a)\mathbf{P}_x(a) = \mathbf{P}_z(a, b).$$

For $\mathbf{P}_x(a) = 0$ we have: $\nu_N(a, b; \phi z) \leq \nu_N(a; \phi z) = \nu_N(a; \phi^{(1)}x)$. But

$$\lim_{N\to\infty} \nu_N(a; \phi^{(1)}x) = \lim_{N\to\infty} \nu_N(a; x) = \mathbf{P}_x(a) = 0.$$

Thus $\mathbf{P}_{\phi z}(a, b) = 0 = \mathbf{P}_z(a, b)$.
2) Let z be the G-collective. Then we obtain

$$\mathbf{P}_{\phi^{(1)}x}(a) = \sum_{b\in L_y} \mathbf{P}_{\phi z}(a, b) = \sum_{b\in L_y} \mathbf{P}_z(a, b) = \mathbf{P}_x(a).$$

In the same way we obtain that $\mathbf{P}_{\phi^{(2)}y}(b) = \mathbf{P}_y(b)$. Finally, we have

$$\mathbf{P}_{\phi z}(a/b) = \mathbf{P}_{\phi z}(a, b)/\mathbf{P}_{\phi^{(1)}x}(a) = \mathbf{P}_z(a, b)/\mathbf{P}_x(a)) = \mathbf{P}_z(b/a),$$

for $\mathbf{P}_x(a) > 0$.

[9]Let ϕ be defined by a function $f = (f_1, f_2(z_1), f_3(z_1, z_2), ...), f_j = 0, 1$. If $f_n(z_1, z_2, ..., z_{n-1}) = 1$, then the element z_n is chosen for a new sequence. Let $x = (x_1, x_2, ..., x_n, ...)$ be a sequence and let $y = (x_m, x_{m+1}, ...), m > 1$. Here $z = (z_j)$ has the form: $z_1 = (x_1, x_m), z_2 = (x_2, x_{m+1}), ...$. Thus, in particular, $f_2(z_1)$ depends (in general) not only on x_1 but also on x_m. Therefore $\phi^{(1)}$ is not a place selection for x.

The reader can easily define for the collectives x and y the relations $\overleftarrow{yx}$, $\overleftarrow{xy}$ and $\overline{xy}$ with respect to the G which are not based, respectively, on mod $\mathbf{P}_{\mathrm{x}}$, mod $\mathbf{P}_{\mathrm{y}}$ and mod $\mathbf{P}$ factorizations. In general $\overleftarrow{yx}$ does not imply $\overleftarrow{xy}$ or that the sequence z is the G-collective (and vice versa).

Probabilities $\mathbf{P}_z(a/b)$, $\mathbf{P}_z(b/a)$, $a \in L_x$, $b \in L_y$, have the meaning of conditional probabilities. If $\mathcal{S}_x$ and $\mathcal{S}_y$ are statistical experiments which generate x and y, respectively, then, for example, $\mathbf{P}_z(b/a)$ is nothing other than the conditional probability of the result b for $\mathcal{S}_y$ if we knew that a was the result of $\mathcal{S}_x$. It is easy to show that probabilities $\mathbf{P}_z(a,b)$ (or $\mathbf{P}_z(a/b)$) can be obtained on the basis of the general definition of conditional probabilities based on the operation of partition. For each $a \in L_x$, we consider the set $A_a = \{u = (a,b),\ b \in L_y\} \subset L_x \times L_y$ and the point set $B_{a,b} = (a,b) \subset A_a$. Let z be an S-sequence (in particular, a collective). It easy to see that the conditional probability $\mathbf{P}(B_{a,b}/A_a)$ (for $\mathbf{P}(A_a) > 0$) defined on the basis of the operation of partition for z coincides with the probability $\mathbf{P}_z(b/a)$. However, the approach based on the operation of combining seems more attractive than the approach based on the operation of partition. In the first case the conditional probabilities have the natural interpretation as a measure of dependence between collectives x and y.

10 Independence of collectives

Let x and y be S-sequences and let $\overleftarrow{yx}$. The y is said to be independent from x if all S-sequences $y(a)$, $a \in L_x$, have the same probability distribution which coincides with the probability distribution $\mathbf{P}_y$ of y. This implies that

$$\mathbf{P}_z(b/a) = \lim_{N\to\infty} \nu_N(b/a;z) = \lim_{N\to\infty} \nu_N(b;y(a))$$

Hence

$$\mathbf{P}_z(a,b) = \mathbf{P}_x(a)\mathbf{P}_y(b),\ a \in L_x,\ b \in L_y. \tag{10.1}$$

Thus the independence implies the factorization of the two dimensional probability $\mathbf{P}_z(a,b)$. However, in general the multiplication rule (10.1) does not imply independence. If (10.1) holds, but $\mathbf{P}_x(a) = 0$, then in principle $\mathbf{P}_z(b/a)$ may depend on a (or it may be that $\mathbf{P}_z(b/a) = \mathrm{Const} \neq \mathbf{P}_y(b)$). By similar reasons the condition "y is independent from x" does not imply that x is independent from y. Dependence on a such that $\mathbf{P}_x(a) = 0$ (or b, $\mathbf{P}_y(b) = 0$) does

not play any role in probabilistic considerations [10]. Therefore it is natural to consider (mod $\mathbf{P}$)- independence.

Let x and y be two (mod $\mathbf{P}$)-combinable S-sequences (or collectives). They are said to be (mod $\mathbf{P}$)-independent if (a) $\mathbf{P}_{y(a)} \equiv \mathbf{P}_y$ for all $a \in L_x$, $\mathbf{P}_x(a) > 0$ and (b) $\mathbf{P}_{x(b)} \equiv \mathbf{P}_x$ for all $b \in L_y$, $\mathbf{P}_y(b) > 0$. In fact, (a) implies (b) and vice versa. For instance, let (a) take place. Then for $\mathbf{P}_y(b) > 0$, we have

$$\mathbf{P}_z(a/b) = \mathbf{P}_z(a,b)/\mathbf{P}_y(b) = \mathbf{P}_z(b/a)\mathbf{P}_x(a)/\mathbf{P}_y(b)$$

$$= \mathbf{P}_{y(a)}(b)\mathbf{P}_x(a)/\mathbf{P}_y(b) = \mathbf{P}_x(a).$$

It is evident that the multiplication rule (10.1) holds for (mod $\mathbf{P}$)-independent sequences. On the other hand, if $\overline{xy}$(mod $\mathbf{P}$) and the multiplication rule (10.1) hold, then x and y are (mod $\mathbf{P}$)-independent.

Remark 10.1. The reader can easily generalize the frequency approach to conditional probabilities and independence to countable sets of labels. Non-countable sets of labels were considered in [102].

Example 10.1. Assume that two coins are tossed simultaneously, the corresponding sequences being x and y (with $L_x = L_y = \{0,1\}$). Our experience says that in mathematical models we can assume that x and y are collectives with probability distributions $\mathbf{P}_x(a)$, $\mathbf{P}_y(b)$, $a, b = 0, 1$. We choose two subsequences of y: (1) $y(0) = (y_{j_l})$, where, for the first coin, $x_{j_l} = 0$; (2) $y(1) = (y_{j_l})$, where, for the first coin, $x_{j_l} = 1$. Our experience says that in the mathematical model (for ordinary coins) we can assume that there exist $\mathbf{P}(b/0) = \lim_{N\to+\infty} \nu_N(b; y(0))$ and $\mathbf{P}(b/1) = \lim_{N\to\infty} \nu_N(b; y(1))$, $b = 0, 1$. If tossing of the second coin does not depend in any way on the tossing of the first coin, then the relative frequencies in $y(0)$ and $y(1)$ have the same behaviour as relative frequencies in y (this is again the experimental fact). Thus we can assume that collectives x and y are independent.

Example 10.2. Assume that an urn contains balls each marked with a number a, where a belongs to the set $S = \{a_1, \dots, a_n\}$. The sequence x is induced by the experiment $\mathcal{S}_x$: we draw a ball from the urn, write its label and return it into the urn. The sequence y is induced by the experiment $\mathcal{S}_y$: after drawing the first ball and before returning it, a second ball is drawn from the urn and its label is written. As usual, we define subsequences $y(a_j)$, $j = 1, \dots, n$, of y. Our experience says that in the mathematical model we can assume that x, y and $y(a_j)$ are collectives and x and y are combinable. Thus the conditional probabilities

[10] Of course, we could not completely exclude the possibility that there may exist physical phenomena in that the dependence on labels having zero probabilities plays some rule.

$\mathbf{P}(b_j/a_i) = \mathbf{P}_{y(a_i)}(b_j)$ are well defined. However, if the distribution of balls in the urn is not symmetric, then $\mathbf{P}(b_j/a)$ depends on a. Thus the collectives x and y are not independent.

11 Frequency and measure - theoretical viewpoints on independence

If we use the frequency approach and take combining as our starting point, then the mathematical and physical conditions for independence concern the interconnection of the two one-dimensional collectives x and y (two statistical experiments $\mathcal{S}_x$ and $\mathcal{S}_y$) or in terms of $\mathbf{P}_z(a, b)$, the type of this two-dimensional distribution: factorization (10.1); they are not concerned with properties on each single collective (statistical experiment). On the other hand, measure-theoretical definition (4.1), (4.2) of independence does not relate in any way to two-dimentional distribution. Of course, definition (4.1), (4.2) can be considered as a generalization of the factorization rule (10.1). However, (4.1), (4.2) extends (10.1) too much. In general (4.1), (4.2) has no relation with original physical motivations of independence. We wish to consider this problem carefully. Consider the following example [88]: The label space S consists of six points $1, \ldots, 6$ with distribution p_i, $i = 1, \ldots, 6$; the event (or set) A consists of the three points $2, 3, 4$, the event C of the two points $1, 2$; the intersection $A \cap C$ is the point 2, and $\mathbf{P}(C/A) = p_2/(p_2 + p_3 + p_4)$ (due to measure-theoretical definition (2.4) or frequency definition (3.4) which is based on the operation of partition). Now the following question is asked: Under what conditions is $\mathbf{P}(C/A)$ equal to $\mathbf{P}(C)$ (or $\mathbf{P}(C \cap A) = \mathbf{P}(A)\mathbf{P}(C)$)? In our example $p_2/(p_2+p_3+p_4) = p_1+p_2$? The example is so chosen that this is true for $p_i = 1/6$, $i = 1, \ldots, 6$. The statement is then made that, in this case, the events A and C are independent. Let us analyze this statement.

Let us consider a set A consisting of the points $2, 3, 4$ and a set C of point 2; here $C \subset A$. Then $\mathbf{P}(2/A) = p_2/(p_2+p_3+p_4)$. Here $\mathbf{P}(2/A)$ certainly does not remain unchanged if we vary the set A, and certainty for no A, $\mathbf{P}(A) \neq 1$, is $\mathbf{P}(2/A)$ equal to p_2. Now, however, in order to make such an equality possible, one consider other sets, C, such that $A \cap C = \{2\}$ but $C \supset A \cap C$. Such subsets of S are, for example, $C_1 = \{1, 2\}$, $C_2 = \{2, 5\}$, $C_3 = \{1, 2, 5, 6\}$, $C_4 = \{1, 2, 5\}$. Then, for each of these C_i, $\mathbf{P}(C_i/A) = p_2/(p_2+p_3+p_4)$. Thus,

having the choice of sets C_i one may ask whether for one or more of them, and with some given distribution, $\mathbf{P}(C/A) = \mathbf{P}(C)$. If all $p_i = 1/6$, this holds true for $C_1 = \{1,2\}$ or for $C_2 = \{2,5\}$ but not for C_3 or C_4. If we take $p_1 = p_5 = 1/12$, $p_2 = p_3 = p_4 = 1/6$, $p_6 = 1/3$, then the above equality holds for $C_4 = \{1,2,5\}$ but no longer for C_1 and C_2, and so on. It seems that the measure-theoretical definition allows the possibility to purely numerical accidents. From the physical point of view, it is not clear: What is the meaning of the statement that, for a given distribution "the events $A = \{2,3,4\}$ and $C = \{2,5\}$ are independent" while "the events $\{2,3,4\}$ and $\{1,2,5\}$ are dependent" or "events $\{1,6\}$ and $\{2,3,4\}$ are dependent"?

One may say that the intersection of two sets A and C has the 'property' of belonging to A and the 'property' of belonging to C (and many others). Nevertheless, the label "2"- the result of the ordinary tossing of one die - is not a two-dimensional label like "blond hair, blue eyes" or "first die 3, second die 5". Therefore a concept of independence of two 'properties' which may or may not influence each other is meaningful. However, this concept must be discussed on the basis of the procedure of combining of collectives corresponding to measurements of these properties.

Conclusion. *Independence should be defined for collectives rather than for isolated events.*

12 Generalization of the operation of combining

In fact, to consider the relation of combining $\widetilde{xy}$ we need not start with two collectives (or S-sequences) x and y. It suffices to have one collective x and a family $\{y(a)\}_{a\in L_x}$ of collectives having the same label set L_y. We denote the system $(x, \{y(a)\}_{a\in L_x})$ by the symbol U_{xy}. In this framework we can also define conditional probabilities $\mathbf{P}_{U_{xy}}(b/a) = \mathbf{P}_{y(a)}(b)$, $a \in L_x$, $b \in L_y$, and two-dimensional probability distribution

$$\mathbf{P}_{U_{xy}}(a,b) = \mathbf{P}_{U_{xy}}(b/a)\mathbf{P}_x(a),\ a \in L_x,\ b \in L_y.$$

In fact, a sequence $z = (z_j)$ corresponding to measurements of pears $z_j = (x_j, y_j)$ may be not defined. Such a situation is common for measurements of so called incompatible observables in quantum mechanics (i.e., observables represented by noncommuting operators), see Chapter 2. In that case it is

impossible to perform a simultaneous measurement of two observables x and y (i.e., we could not form the collective z). Nevertheless, we could speak about two properties A and B of the physical system. The conditional probability $\mathbf{P}_{U_{xy}}(b/a)$ has the following meaning: if the result of a measurement of the property A is equal a, then the probability to obtain the value b of the property B is equal $\mathbf{P}_{U_{xy}}(b/a)$.

We suppose now that it is also possible to perform a measurement of the property B and, for each $B = b$, to perform a measurement of the property A. Mathematically such measurements are described by a collective y (corresponding to a measurement of B) and a system $\{x(b)\}_{b \in L_y}$ of collectives (corresponding to measurements of A under the condition $B = b$). Thus we have the system $U_{yx} = (y, \{x(b)\}_{b \in L_y})$. Here we can also define the conditional probabilities $\mathbf{P}_{U_{yx}}(b/a)$ and two-dimensional probability distribution

$$\mathbf{P}_{U_{yx}}(b, a) = \mathbf{P}_{U_{yx}}(a/b)\mathbf{P}_y(b), \; a \in L_x, \; b \in L_y.$$

It may be that $\mathbf{P}_{U_{xy}}(a, b) \neq \mathbf{P}_{U_{yx}}(b, a)$. In such a case the two-dimensional probability distribution $\mathbf{P}(a, b)$ corresponding to pears $(A = a, B = b)$ does not exist.

13 Comparative probability

All probability models discussed in the previous sections are called quantitative probability models. Terse the quantitative statement "$P(A) = p$" read "the probability of A is p" is the basis of these theories[11]. On the other hand, the modal or classificatory statement "A is probable" or "A is likely" seems to be most common in ordinary discourse. To formalize such an approach, we can consider, for example, a binary relation P_2 in the set $D \times D$, where D is the set of events. This relation can be read as follows: If $(A, B) \in P_2$, then A is at least as probable as B, $A \geq B$. Such a formalization gives so called *comparative probability* formalism (see, for example, T. Fine [39]).

Comparative probability induces more extended class of probability models (with larger domains of application) than quantitative probability.

For example, having observed that 10 tosses of a strange coin resulted in 7 heads, we are more justified in asserting that "heads are more probable than tails" then asserting that "the probability of heads is 0.7". There

[11] There arises natural question Why do we consider only real numbers p as quantities?

exist relatively simple mathematical models in that we consider to be valid comparative probability statements that, are incompatible with any representation in quantitative theory[12].

However, my opinion is that comparative probability models have to be considered as "derivatives" of the three fundamental models (classical, frequency and ensemble). To define the binary relation P_2, we need to use one of fundamental models (or their generalizations).

Typically it is assumed that the binary relation P_2 satisfies the following axioms (see T. Fine [39], p.17):

C0. (Nontriviality) $\Omega \succ \emptyset$, where $\emptyset$ is the null or empty set.

C1. (Comparability) $A \succsim B$ or $B \succsim A$.

C2. (Transitivity) $A \succsim B$, $B \succsim C \Rightarrow A \succsim C$.

C3. (Improbability of impossibility) $A \succsim \emptyset$.

C4. (Disjoint unions) $A \cap (B \cup C) = \emptyset \Rightarrow (B \succsim C \Leftrightarrow A \cup B \succsim A \cup C)$.

Axiom C1 and C2 establish that the relation $\succsim$ is a linear complete order, The requirement that all events be comparable is not insignificant and as been denied by some authors [39]. To illustrate the latter possibility, we consider the following example. There is an ensemble $S, |S| = N$, of coins having different centers of mass. The first coin tossing experiment (for all coins $s \in S$) gave N_1 heads and the second experiment give N_2 heads. If $N_1 > M_1 = N - N_1$, but $N_2 < M_2 = N - N_2$, then we cannot assert neither "heads are least as probable as tails" $(A \succsim B)$ nor "tails are at least as probable as heads".

In Chapter 4 we construct a quantitative probability model (with the field of p-adic numbers as quantitative space) that induces a comparative probability model in that there exist noncomparable events. In this model the axiom (C4) is also violated.

Remark 13.1. (Subjective probability as comparative probability). It seems that the comparative interpretation is the one of possible interpretation of subjective probability. We remark that is does not sound reasonable to use the fixed ordered set (the segment $[0, 1]$) for quantitative representation of subjective probabilities. The use of $[0, 1]$ is the root of misunderstandings

[12] At least if $\mathbf{R}$ is used as a "quantitative space".

related to subjective probability. This implies that numbers $p \in [0, 1]$ are often interpreted as frequency of ensemble probabilities.

Chapter 2

Quantum probabilities

In this Chapter probability interpretations of quantum mechanics will be discussed. The main attention will be paid to comparative analysis of statistical measurements for classical and quantum physical systems. We shall see that there is (at least formally) a difference in statistical behaviors of classical and quantum physical systems. We present a purely probabilistic explanation of this difference.

1 Classical and quantum probability rules

1. Properties of physical systems. The notion of a *property* of a physical system[1] will play the important role in our analysis. Therefore we shall discuss this notion. In this discussion we shall use not only physical but also philosophical arguments. The reader who is not interested in such general discussions may start directly with a probability interpretation of quantum states (see subsection 2).

Before to start the discussion, we consider following simple examples of physical properties.

Example 1.1. Let S be an ensemble of rigid bodies. Suppose that these bodies have one of two colours, black or white, and one of two forms, ball or cube. The colour and form are properties of $s \in S$. Numerically these

[1] In fact, the notion of a property of a physical system is (more or less) equivalent to the standard notion of a physical observable. However, an observable must be connected with some observation (measurement). On the other hand, we wish to consider properties of physical systems which in principle are not related to measurements.

properties, A and B, can be described by quantities $A = 0, 1$ for black and white bodies, respectively, and $B = 0, 1$ for ball and cube, respectively.

Example 1.2. Let a be a particle (classical or quantum). The position $\mathbf{q}$ and momentum $\mathbf{p}$ of a are properties of a. Numerically these properties are described by continuous spectrum of values (by the field of real numbers $\mathbf{R}$). In what follows we shall mainly study properties which are described by discrete spectra of values. In the case of the position and momentum we can make the following discretization. Let D_q and D_p be domains in $\mathbf{R}^3$. We set $A = 1$ if $\mathbf{q} \in D_q$ and $A = 0$ if $\mathbf{q} \notin D_q$; $B = 1$ if $\mathbf{p} \in D_p$ and $B = 0$ if $\mathbf{p} \notin D_p$. The quantities A and B are properties of a. In fact, the following question will be studied in this section: What is a difference in statistical behaviours of the properties A and B for classical and quantum particles?

In the physical community there is no the unique point of view to the notion of a property of a quantum physical system (see, for example, [36], [43], [109], [7], [8], [10]–[12], [13], [14], [27], [28], [33] [47], [24], [46], [29], [30], [81] for the details). Some scientists keep to **realism.** They assume that a property is an *objective* characteristic of a quantum system. Thus a property is a property of the object. Such a property is not related to the act of a measurement. It does not depend on a subjective observer. In particular, adherents of realism (L. de Broglie, A. Einstein, D. Bohm,...) think that both classical and quantum particles have well defined (localized) positions and momentums. Adherents of realism can be split in two subgroups. This splitting is based on two different viewpoints to the following problem. Does the quantum formalism operate with *initial* values of properties (i.e., values before acts of measurements) or *final* values of properties (i.e., values after acts of measurements)? This problem is very important in quantum mechanics, because here a measurement can change crucially values of properties of physical systems. We shall call adherents of the initial values hypothesis ***i*-realists** and adherents of the final values hypothesis ***f*-realists**[2].

Other part of the physical community supports the ideas of **empiricism.** They assume that a quantum formalism does not describe microreality as such. Properties obtained via quantum measurements are not properties of quantum systems (not properties of the object). They are merely properties of measurement phenomena (properties of instruments and physical

[2]For example, any measurement of the position of a quantum particle should change the localization of this particle. *I*-realists suppose that a quantum measurement gives the initial value of the position, $q = q_i$; f-realists suppose that a quantum measurement gives the final value of the position, $q = q_f$.

circumstances in that these instruments are used). In particular, adherents of empiricism (N. Bohr, W. Heisenberg, J. von Neumann,...) claim that positions and momentums of quantum particles are not objective. For example, an electron has no definite position before the act of a measurement. It is just a 'cloud' which is *collapsed* into a point by the act of a measurement. Many adherents of empiricism think that a property depends not only on a measurement procedure $\mathcal{M}$, but also on so called preparation procedure $\mathcal{E}$ (see section 2 for the details) which is used to prepare quantum systems for acts of measurements. We call the adherents of the latter viewpoint $\mathcal{E}/\mathcal{M}$-empiricists.

Example 1.3. (Nonobjective property) Let us keep to $\mathcal{E}/\mathcal{M}$-empiricism. As cats cannot fly, the speed of flying v cannot be considered as an objective property of the cat. We consider the following procedure $\mathcal{E}$ to prepare cats to fly: each cat is located near the desk of an airplane which has a system of the autopilot; by manipulations with buttons a cat can change the speed v of the airplane. A large statistical ensemble S of cats in airplanes is prepared by $\mathcal{E}$. A measurement procedure $\mathcal{M}$ is a measurement of the speed v of an airplane with the cat. The v is not a property of the cat (on the other hand, it is not just a property of the airplane). It is merely a property of the preparation procedure $\mathcal{E}$. If cats can choose only a finite set of speeds $v_1, ..., v_k$, then the measurement $\mathcal{M}$ will produce discrete probability distribution $\mathbf{P}(v = v_i), i = 1, 2, ..., k$.

Empiricism is often identified with **idealism.** By idealists viewpoint quantum systems have no objective properties at all. This approach immediately implies a *death of reality* (not only reality of the microworld, but also reality of macroworld which is composed of microsystems). However, in principle empiricism need not imply idealism. It is very well possible to believe in the objective existence of atoms and electrons without being committed to the thesis that this reality is described by the quantum mechanical formalism.

The realist philosophy is very attractive for scientists working in classical physics. However, we shall see that the realist viewpoint induces some problems (so called Einstein-Podolsky-Rosen paradox) in the foundations of quantum physics[3]. The empiricists approach seems to be free of such problems. However, empiricism is not so attractive as the philosophic basis for the

[3]In fact, we shall show that these problems have purely mathematical origin and they are connected only with the foundations of probability theory.

investigation of reality. If we even do not keep to idealism and not deny existence of objective reality (which is independent to our observations), then by the empiricists ideology we still have to assume that the quantum formalism describes not objective reality of microworld, but only reality of equipment in our laboratories.

Of course, such a classification of members of the physical community is not rigid. Some of them balance on the boundary between different viewpoints.

We shall see that different viewpoints to the notion of a property imply different viewpoints to the notion of probability. For example, i-realists can use both ensemble and frequency probability formalisms; f-realists and empiricists can use only frequency probability formalism.

2. Probability interpretation of a quantum state. We discuss now a probability interpretation of quantum mechanics. We may restrict our considerations to two dimensional quantum systems. Already such quantum systems demonstrate all delicate features of this problem. Let us consider a large statistical ensemble S of quantum systems[4] s having two properties, A and B. Let $\mathcal{H} = \mathbf{C} \times \mathbf{C}$ be the two dimensional complex linear space with the inner product $(\cdot, \cdot) : \mathcal{H} \times \mathcal{H} \to \mathbf{C}$. In the quantum formalism the statistical ensemble S is described by a normalized vector $\phi \in \mathcal{H}$ (i.e., $\|\phi\|^2 = (\phi, \phi) = 1$). This vector is called a *quantum state.* The properties (physical observables) A and B are described by symmetric operators $\hat{A}$ and $\hat{B}$, respectively. Let $e_A = (\phi_0, \phi_1)$ and $e_B = (\psi_0, \psi_1)$ be two orthonormal bases in $\mathcal{H}$ consisting of eigenvectors of the operators $\hat{A}$ and $\hat{B}$, respectively. The quantum state ϕ can be represented in two ways:

$$\phi = c_0\phi_0 + c_1\phi_1, \text{ where } c_0, c_1 \in \mathbf{C},\ |c_0|^2 + |c_1|^2 = 1; \tag{1.1}$$

$$\phi = d_0\psi_0 + d_1\psi_1, \text{ where } d_0, d_1 \in \mathbf{C},\ |d_0|^2 + |d_1|^2 = 1. \tag{1.2}$$

By the probability interpretation of expansion (1.1) of the quantum state ϕ the probability $\mathbf{P}(A = \alpha)(\equiv \mathbf{P}_\phi(A = \alpha))$ that $s \in S$ has the property $A = \alpha$ is equal to $|c_\alpha|^2$. In the same way expansion (1.2) gives that $\mathbf{P}(B = \beta)(\equiv \mathbf{P}_\phi(B = \beta)) = |d_\beta|^2$. The possibility to expand one basis with respect

[4]Quantum theory is not yet understood as well as e.g. classical mechanics or special relativity. In particular, there is no precise definition of a quantum system. The unique way to extract 'quantum domain' from the classical world is to use statistical properties of quantum systems (which are, in fact, defined via these statistical properties).

to other basis induces connection between the probabilities $\mathbf{P}(A = \alpha)$ and $\mathbf{P}(B = \beta)$. Let us expend the vectors ϕ_0 and ϕ_1 with respect to the basis e_B :

$$\phi_0 = u_{00}\psi_0 + u_{01}\psi_1, \text{ where } u_{0\alpha} \in \mathbf{C}, \ |u_{00}|^2 + |u_{01}|^2 = 1; \tag{1.3}$$

$$\phi_1 = u_{10}\psi_0 + u_{11}\psi_1, \text{ where } u_{1\alpha} \in \mathbf{C}, \ |u_{10}|^2 + |u_{11}|^2 = 1. \tag{1.4}$$

Thus $d_0 = c_0u_{00} + c_1u_{10}$, $d_1 = c_0u_{01} + c_1u_{11}$ and we obtain the *quantum rule* for transformation of probabilities:

$$\mathbf{P}(B = \beta) = |c_0u_{0\beta} + c_1u_{1\beta}|^2, \ \beta = 0, 1. \tag{1.5}$$

3. Contradiction. On the other hand, by the probability interpretation of expansions (1.3), (1.4) we obtain that $\mathbf{P}(B = \beta/A = \alpha) = |u_{\alpha\beta}|^2$. Indeed, in (1.3), (1.4) the quantum states $\phi_\alpha, \alpha = 0, 1$, describe statistical ensembles $\bar{S}(A = \alpha)$ of physical systems which have the property $A = \alpha$. Therefore the expansion of the ϕ_α with respect to the basis e_B gives corresponding probabilities for $B = \beta$ (under the condition that $A = \alpha$). By the formula of total probability we obtain:

$$\mathbf{P}(B = \beta) = \sum_{\alpha=0,1} \mathbf{P}(A = \alpha)\mathbf{P}(B = \beta/A = \alpha) = |c_0|^2|u_{0\beta}|^2 + |c_1|^2|u_{1\beta}|^2. \tag{1.6}$$

Thus in general 'quantum rule' (1.5) differs from 'classical rule' (1.6). The standard viewpoint to the contradiction between (1.5) and (1.6) is that this is the exhibition of the *non-Boolean structure* (violation of Bayes' formula) of quantum statistical ensembles. It is typically pointed out that this non-Boolean structure implies that it is impossible to consider a wave function ϕ as the description of a statistical ensemble S of identically prepared objects (not interacting with any prepared or measuring instruments). However, careful analysis will show that this contradiction is a consequence of formal manipulations with Kolmogorov probabilities. On one hand, this contradiction need not imply the non-Boolean structure of quantum statistical ensembles and in principle we need not use a new 'quantum probabilistic calculus' for the description of quantum phenomena [5]. On the other hand,

[5]One of the main aims of this book is to demonstrate that, in fact, there is no difference between 'quantum and classical probabilities'. Behaviour of probabilities related to so called quantum systems can be explained on the basis of classical probability theory (which is, of course, not reduced to Kolmogorov's measure-theoretical formalism). As

we shall see that even non-Boolean structure (in fact, nonexistence of conditional probabilities and consequently the impossibility to use Bayes' formula or the formula of total probability) need not imply that a statistical ensemble should have some 'special quantum structure'. Such a non-Boolean structure can be easily found in classical statistical phenomena.

Remark 1.1. The reader can easily understand that (1.5) differs from (1.6) only if the operators $\hat{A}$ and $\hat{B}$ do not commute. Observables (properties) A, B which are represented by noncommuting operators $\hat{A}, \hat{B}$ are said to be *incompatible.* By the quantum formalism incompatible observables cannot be measured simultaneously. Hence, for incompatible, observables A, B the two dimensional probability distribution $\mathbf{P}(\alpha, \beta) = \mathbf{P}(A = \alpha, B = \beta)$ cannot be defined on the basis of real physical measurements.

To find roots of the contradiction between (1.5), (1.6), we have to analyse the meaning of probabilities in formulas (1.5), (1.6). We shall do this in section 3.

2 Physical interpretations of the wave function

There are many different physical interpretations of the wave function[6] ϕ. We discuss the most important interpretations.

1. Ensemble realist interpretation. Typically this interpretation is called a **statistical interpretation** (following to L. Ballentine, [8]). Here it is assumed that ϕ describes a statistical ensemble S of identically prepared objects s. Properties of s are their objective properties. On the basis of this interpretation, it is possible to keep to both i-realism or f-realism. However, the main part of investigations on the basis of the statistical interpretation is based on i-realism. It seems that, for example, A. Einstein was (more or less) an adherent of the statistical interpretation (in the i-realists framework).

the operational definition of a quantum system is given by the description of statistical behaviour, the realization of such a program will imply that there is no crucial difference between classical and quantum physics. Therefore not only quantum systems may exhibit some 'classical probabilistic features', but also classical systems may exhibit some 'quantum probabilistic features'. In particular, quantum probabilistic behaviour can be exhibited by macrosystems.

[6] Of course, such a proliferation of paradigms is characteristic of a crisis in the development of quantum theory, see [30], [33] for the details.

Here a wave function ϕ describes probabilistic distributions of properties of elements s of a statistical ensemble S. If we keep to i-realism, then these are distributions of initial properties of these elements; if we keep to f-realism, then these are distributions of final properties (obtained via measurement) of these elements.

2. Individual realists interpretation. Here it is assumed that ϕ describes not a statistical ensemble S, but an individual physical system s. Properties of s are considered as being objective.

3. Ensemble empiricists interpretation. Here it is assumed that ϕ describes a statistical ensemble which does not consist of quantum systems, but it merely consists of preparation and measurement procedures (see section 2) in that these quantum systems are involved. Probabilistic distributions which are related to this statistical ensemble are merely (see Example 1.3) distributions of these preparation and measurement procedures. This interpretation is typically considered as a **conventional interpretation** of quantum mechanics.

4. Individual empiricists interpretation. Here it is assumed that ϕ describes an individual quantum system s. Thus ϕ contains information on probabilistic distributions of possible reactions of s to acts of measurements. Many adherents of this interpretation keep to idealism and suppose that s has no objective properties at all. 'Properties' of s are connected only with specific acts of measurements. This extreme form of the individual empiricists interpretation is called an **orthodox Copenhagen interpretation.** Typically N. Bohr is mentioned as one of the main adherents of the orthodox Copenhagen interpretation. However, it seems that N. Bohr has no idealists views to quantum reality. He was merely an adherent of empiricism (and he balanced between the individual and ensemble interpretations).

5. Individual interpretation and probability. We shall not consider in this book individual interpretations of ϕ. It seems that the use of subjective probability is the only reasonable way to give the probabilistic foundation for these interpretations. However, it would be rather strange to use such an argument as a 'measure of the personal belief' as the cornerstone of the fundamental physical theory. As it has been mentioned in section 7, Chapter 1, at the moment the only real possibility to justify the use of subjective probabilities is to reduce them to ensemble or frequency probabilities. Let ϕ describe the measure of our belief that, for example, the position $\mathbf{q}$ of an (individual) electron would be observed in a domain D. But how can this measure of our belief be found? The only way is to use our ensemble or

frequency experience[7].

Therefore we shall pay our attention to ensemble interpretations. These interpretations can be used on the basis of ensemble and frequency probability theories. Ensemble probability theory provides the basis of the statistical interpretation in the framework of i-realism. Frequency probability theory must be used as the basis of the statistical interpretation in the f-realists framework and the ensemble empiricists interpretation.

Our main purpose is to demonstrate that the contradiction between the quantum and classical probabilistic rules (which is often reduced to the non-Boolean structure of quantum statistical ensembles) cannot be used as a reason in favour or against some interpretation of quantum mechanics. Therefore physicists have to solve their own problems and find new physical reasons to choose the right physical interpretation of the wave function.

6. Preparation and measurement procedures. Many physicists imagine a quantum measurement process as it is split in two procedures: 1) a *preparation procedure* $\mathcal{E}$, 2) a *measurement procedure* $\mathcal{M}$.

A preparation procedure $\mathcal{E}$ produces an ensemble S of quantum systems s having the definite probability distribution of some property A. This property can be considered as an objective property of quantum systems $s \in S$ (realism) or as merely a property of the preparation procedure $\mathcal{E}$ (empiricism). In the later case probabilities $\mathbf{P}(A = \alpha)$ are fixed by $\mathcal{E}$. This statistical information is algebraically coded in expansion (1.1) of the quantum state ϕ which represents the ensemble S.

Probabilities $\mathbf{P}(A = \alpha)$ have different meanings in different approaches to quantum mechanics. By the statistical (ensemble realists) interpretation in the i-realists framework $\mathbf{P}(A = \alpha) = \mathbf{P}_S(A_i = \alpha)$ is the distribution of initial values A_i of the property A in the ensemble S. By the same statistical interpretation, but in the f-realists framework $\mathbf{P}(A = \alpha) = \mathbf{P}_a(A_f = \alpha)$ is the distribution of final values A_f of the property A in the collective $a = (\alpha_1, \alpha_2, ..., \alpha_k, ...)$ induced by measurements $\mathcal{N}$ of A for $s \in S$. The measurement $\mathcal{N}$ is not performed, because it would disturb the result of the preparation procedure $\mathcal{E}$. However, in principle it can be done (to test functioning of the preparation procedure $\mathcal{E}$). If we keep to the ensemble empiricists interpretation, then probabilities $\mathbf{P}(A = \alpha)$ have also the

[7]Another possibility to deal with subjective probabilities is to consider them as measures of information. However, even if probability is considered as subjective information, then it can be again reduced to ensemble or frequency probability for distribution of ideas in the brain.

frequency meaning. The only difference is that by the latter interpretation values of A are considered as properties of the preparation procedure $\mathcal{E}$: $\mathbf{P}(A = \alpha) = \mathbf{P}_a(A_{\mathcal{E}} = \alpha)$.

Typically the $\mathcal{E}$ is realized as a filter with respect to values of A. For example, to prepare the statistical ensembles $\bar{S}(A = \alpha), \alpha = 0, 1$, (described by pure states ϕ_α) we can use filters F_α corresponding to fixed values[8] α of A.

The next step is a measurement procedure $\mathcal{M}$. The $\mathcal{M}$ is used for the measurement of other property[9] B for elements s of the ensemble S which has been prepared by $\mathcal{E}$ on the basis of the property A. The main feature of the quantum mechanical formalism is that theoretically the probability distribution for B can be found on the basis of the purely algebraic computations via representation (1.2).

Remark 2.1. The formalism of quantum mechanics is used in the following way. The preparation procedure $\mathcal{E}$ fixes probabilities $\mathbf{P}_S(A = \alpha), \alpha = 0, 1$. Thus we can found the absolute values of coefficients c_α in expansion (1.1) which determine the quantum state ϕ representing the ensemble S. On the other hand, the probability formalism does not determine the phases θ_α of the coefficients $c_\alpha = e^{i\theta_\alpha}\sqrt{\mathbf{P}_S(A = \alpha)}$. These phases must be found on the basis of other physical reasons. Then we can expand ϕ with respect to other basis $e_B = (\psi_0, \psi_1)$ corresponding to fixed values of the property B and find coefficients d_β of this expansion. The quantities $|d_\beta|^2$ give theoretical values for probabilities $\mathbf{P}(B = \beta)$. These theoretical values are compared with experimental values obtained on the basis of the measurement $\mathcal{M}$ for the observable B.

3 'Contradiction' between quantum and classical probability calculi

In this section we demonstrate that one of possible roots of the contradiction between quantum rule (1.5) (non-Boolean probability theory) and classical rule (1.6) (Boolean probability theory) is the identification of conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ for the quantum state ϕ, see (1.2), with probabilities $\mathbf{P}(B = \beta)$ for the quantum states $\phi_\alpha, \alpha = 0, 1$, see (1.3),

[8]Only particles with the property $A = \alpha$ can pass F_α.

[9]The values of B obtained via $\mathcal{M}$ are considered (depending on a viewpoint) as B_i (objective initial values), B_f (objective final values) or $B_{\mathcal{E}/\mathcal{M}}$ (values determined by the preparation procedure $\mathcal{E}$ and the measurement $\mathcal{M}$).

(1.4). Another possible root is nonexistence of conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$. In the latter case theory is, of course, non-Boolean, but nevertheless it is still 'classical' (in the sense that such behaviour can be easily simulated for macrosystems).

Thus the contradiction between quantum and classical rules (1.5) and (1.6) is a consequence of the formal use of Kolmogorov's axiomatics. As we have already discussed, A.N. Kolmogorov eliminated the concrete structures of probabilistic spaces from his model. Thus there arose rather mystical symbol $\mathbf{P}$ of abstract probability which was not related to any concrete statistical ensemble S or collective x. It seems that in quantum physics Kolmogorov's abstract probabilities are often used formally. This implies identification of (conditional) probabilities which are related to different ensembles or collectives. However, such probabilities need not be equal. On the other hand, Kolmogorov's definition of conditional probabilities via Bayes' formula induces the opinion that (at least in 'classical probability theory') existence of probabilities $\mathbf{P}(E_j), j = 1, 2$, with $\mathbf{P}(E_2) > 0$ must automatically imply existence of conditional probability $\mathbf{P}(E_1/E_2)$. However, such an assumption need not hold true for all statistical phenomena. In the ensemble probability framework (see section 2, Chapter 1) we need not assume that the family of all events $F(\pi_S)$ (determined by the family π_S of properties of elements $s \in S$) is an algebra. Thus $E_1, E_2 \in F(\pi_S)$ need not imply $E_1 \cap E_2 \in F(\pi_S)$. Here conditional probability $\mathbf{P}(E_1/E_2)$ could not be defined by Bayes' formula.

1. Ensemble approach: disturbance effects. Let us we follow i-realism. Both properties A and B are objective properties of elements of the statistical ensemble S represented by the quantum state ϕ. Measurements give initial values of these properties, $A \equiv A_i, B \equiv B_i$. We can consider sub-ensembles $S(A = \alpha)$ and $S(B = \beta), \alpha, \beta = 0, 1$, of S which consist of elements s having the properties $A = \alpha$ and $B = \beta$, respectively. By the ensemble definition of probability $\mathbf{P}_S(A = \alpha) = |S(A = \alpha)|/|S|$ and $\mathbf{P}_S(B = \beta) = |S(B = \beta)|/|S|$ and by the ensemble definition of the conditional probability

$$\mathbf{P}_S(B = \beta/A = \alpha) = \mathbf{P}_{S(A=\alpha)}(B = \beta) = \frac{|S(A = \alpha) \cap S(B = \beta)|}{|S(A = \alpha)|} . \quad (3.1)$$

We can use Bayes' formula (and the formula of total probability) for these probabilities. It seems that we should obtain the above contradiction. However, there is one delicate point.

In general we cannot assume that the conditional probabilities $\mathbf{P}_S(B = \beta/A = \alpha) = \mathbf{P}_{S(A=\alpha)}(B = \beta)$, $\alpha, \beta = 0, 1$, can be obtained from expansions (1.3), (1.4). We cannot identify the sub-ensembles $S(A = \alpha), S(B = \beta)$ of S with ensembles $\bar{S}(A = \alpha), \bar{S}(B = \beta)$ which are described by the quantum states ϕ_α and ϕ_β, respectively.

There are different preparation procedures $\mathcal{E}, \mathcal{E}(A = \alpha), \mathcal{E}(B = \beta)$, $\alpha, \beta = 0, 1$. They produce ensembles $S, \bar{S}(A = \alpha), \bar{S}(B = \beta)$, respectively, which are represented by quantum states $\phi, \phi_\alpha, \psi_\beta$, respectively. We cannot identify the sub-ensembles $S(A = \alpha), S(B = \beta)$ of S with ensembles $\bar{S}(A = \alpha), \bar{S}(B = \beta)$. For instance, the preparation procedures $\mathcal{E}(A = \alpha), \alpha = 0, 1$, can be realized as filters $F(A = \alpha)$ such that only quantum systems s with the property $A = \alpha$ can pass $F(A = \alpha)$. However, such a filtration changes the value of the property B for s. Thus in general

$$\mathbf{P}_S(B = \beta/A = \alpha) = \mathbf{P}_{S(A=\alpha)}(B = \beta) \neq \mathbf{P}_{\bar{S}(A=\alpha)}(B = \beta).$$

Moreover, $\mathbf{P}_S(B = \beta/A = \alpha)$ may be not well defined. Despite of the fact that $A, B \in \pi_S$, it may be that the set $\{A = \alpha\} \cap \{B = \beta\}$ is not described by any property $C \in \pi_S$.

Such a phenomenon is not essentially nonclassical. An example of the selection of a sub-ensemble on the basis of one fixed property which can change the probability distribution of other property can be easily found for classical systems. We can illustrate this problem by Example 1.1. Let S be an ensemble of bodies having different colours, $A = 0, 1$, and forms, $B = 0, 1$. There are sub-ensembles $S(A = \alpha), \alpha = 0, 1$, corresponding to fixed colours and $S(B = \beta), \beta = 0, 1$, corresponding to different forms. To extract elements of the S having the fixed colour α, we use a device D_α which changes randomly the form of a body (some bodies of the form $B = 0$ are transformed in bodies of the form $B = 1$ and vice versa). By this procedure we obtain new ensembles $\bar{S}(A = \alpha), \alpha = 0, 1$. Of course, the distributions of B in $\bar{S}(A = \alpha), \alpha = 0, 1$, may differ from the initial distributions of B in the ensembles $S(A = \alpha), \alpha = 0, 1$.

Conclusion. *The contradiction between 'quantum and classical probabilistic rules' (1.5), (1.6) need not be regarded to the specific ('nonclassical') behaviours of statistical ensembles of quantum systems. The possible root of this contradiction is the formal use of Kolmogorov's measure-theoretical approach in that we do not control the relation between probabilities and statistical ensembles. The identification of probabilities corresponding to different*

statistical ensembles implies (in general) the use of wrong values, $|u_{\alpha\beta}|^2$, for conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ (which, in fact, must be calculated on the basis of (3.1)). This induces the illusion of the violation of Bayes' formula (and the formula of total probability) in the quantum formalism.

2. Ensemble approach: no conditional probabilities. It must be pointed out that any quantum state ϕ represents not a finite statistical ensemble consisting of N quantum systems, but an infinite ideal statistical ensemble S. For any property C, probabilities $\mathbf{P}_\phi(C = \gamma)$ are, in fact, probabilities $\mathbf{P}_S(C = \gamma)$ with respect to this infinite ensemble S. Of course, in each concrete run $R = \{1, 2, ..., N\}$ of experiments we can obtain only a finite statistical ensemble $S^{(R)}$ and 'experimental ensemble probabilities' are probabilities (relative frequencies) with respect to $S^{(R)}$:

$$\mathbf{P}_\phi^{\exp}(C = \gamma) = \mathbf{P}_{S^{(R)}}(C = \gamma) = \frac{|\{s \in S^{(R)} : C = \gamma\}|}{|S^{(R)}|} .$$

For different runs R and R', these probabilities are rather different. The main feature of quantum systems (as many other physical systems) is that these probabilities have the property of the statistical stabilization, namely, $\lim_{|R| \to \infty} \mathbf{P}_{S^{(R)}}(C = \gamma) = |d_\gamma|^2$, where d_γ are coefficients in the expansion of ϕ with respect to the system of eigenvectors of the operator $\hat{C}$. These limiting probabilities are probabilities with respect to the infinite ideal ensemble S :

$$|d_\gamma|^2 = \mathbf{P}_\phi(C = \gamma) = \frac{|\{s \in S : C = \gamma\}|}{|S|} . \tag{3.2}$$

However, as the field of real numbers $\mathbf{R}$ does not contain actual infinities, formula (3.2) has no meaning in the framework of real analysis. Instead of (3.2), mathematicians (and, as a consequence, physicists) use the measure-theoretical approach. However, (despite of the common opinion) this approach cannot be used as a justification of the ensemble probability theory even in the case of a countable ensemble $S = \{s_1, s_2, ..., s_k, ...\}$. For example, let us try to define the uniform σ-additive probability on $S : \mathbf{P}(\{s_1\}) = \mathbf{P}(\{s_2\}) = ... = \mathbf{P}(\{s_k\}) ... \neq 0$. Then $\mathbf{P}(S) = \sum_{j=1}^{\infty} \mathbf{P}(\{s_j\}) = \infty$.

Such 'pathological' properties of the field of real numbers (the absence of actual infinities) is one the reasons to use the ensemble-frequency interpretation instead of the purely ensemble interpretation. By the ensemble-frequency interpretation a Kolmogorov probability space is based not on the ensemble S of quantum systems, but (roughly speaking) on the ensemble Ω

of all possible (ideal infinite) runs of experiments. However, this approach to the definition of a probability space was, in fact, never used by physicists. They typically assume that a Kolmogorov probability space gives the mathematical representation of an ensemble of quantum systems. One of the main reasons to do so and to reject the ensemble-frequency approach is the impossibility to construct a probability measure on the space of all runs for measurements corresponding incompatible properties.

One of problems of the Kolmogorov axiomatics is that probability $\mathbf{P}$ must be closed (defined on the σ-algebra or at least algebra of sets). Thus if probabilities of events $\{A = \alpha\}$ and $\{B = \beta\}$ are well defined, then automatically probability of the event $\{A = \alpha\} \cap \{B = \beta\}$ must be well defined. Hence the conditional probability $\mathbf{P}(B = \beta/A = \alpha)$ must be well defined in Kolmogorov's framework. However, such conditional probabilities are not observed. In principle, there are no reasons to assume that they are even well defined for each quantum state ϕ. For example, why we cannot assume that the ensemble $S = \mathbf{N}$ and 'probability' $\mathbf{P} = \delta$, where δ is the density of natural numbers? In such situations there is no Bayes' formula at all and the problem of difference between 'quantum and classical probability rules' is meaningless. Of course, these are non-Boolean models. However, this non-Boolean structure of probabilities has no special nonclassical features.

Detailed analysis of the problem of existence of conditional probabilities will be presented in section 4.

3. Frequency probability viewpoint to quantum probabilistic rule. The frequency approach to probability gives more freedom than the ensemble approach. Here we need not assume that properties of quantum systems are objective[10]. Thus in principle we can consider various combinations of objective and nonobjective properties of quantum systems.

We start with the general scheme in that we do not suppose that any of properties $A(= 0, 1)$ and $B(= 0, 1)$ has an objective character. Let $\phi \in \mathcal{H}$ be a quantum state. As in section 1, we consider the two dimensional Hilbert space $\mathcal{H}$. Properties A and B are represented by symmetric operators $\hat{A}$ and $\hat{B}$; $e_A = (\phi_0, \phi_1)$ and $e_B = (\psi_0, \psi_1)$ are orthonormal bases in $\mathcal{H}$ consisting of eigenvectors of operators $\hat{A}$ and $\hat{B}$, respectively. Thus we have: $\phi = c_0\phi_0 + c_1\phi_1 = d_0\psi_0 + d_1\psi_1$, where $c_0, c_1, d_0, d_1 \in \mathbf{C}$, and $|c_0|^2 + |c_1|^2 =$

[10]We recall (see section 2) that nonobjective character of some properties (creation of these properties in the process of a measurement) does not imply 'essentially quantum features' of systems.

$1, |d_0|^2 + |d_1|^2 = 1$.

A quantum state ϕ represents an *ideal infinite ensemble* S of quantum systems. This ensemble is characterized in the following way: frequency probability distribution of any property $C(= 0,1)$ is given by squares of absolute values of coefficients in the expansion of ϕ with respect to the system of eigenvalues of the operator $\hat{C}$ representing C. In particular, we have:

(1) Any series of measurements $\mathcal{N}$ of the property A for elements $s_j \in S, j = 1, 2, ...$, induces a collective

$$a_{\mathcal{N}} = (\alpha_1, \alpha_2, ..., \alpha_k, ...),\ \alpha_j = 0, 1,$$

such that frequency probabilities $\mathbf{P}_{a_{\mathcal{N}}}(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a_{\mathcal{N}})$ are equal to $|c_\alpha|^2, \alpha = 0, 1$. Here, as usual, $\nu_k(\alpha; a_{\mathcal{N}}) = n_k(\alpha; a_{\mathcal{N}})/k$ is the relative frequency of realizations of the value $A = \alpha$.

(2) Any series of measurements $\mathcal{M}$ of the property B for elements $s_j \in S, j = 1, 2, ...$, induces a collective

$$b_{\mathcal{M}} = (\beta_1, \beta_2, ..., \beta_k, ...),\ \beta_j = 0, 1, \tag{3.3}$$

such that frequency probabilities $\mathbf{P}_{b_{\mathcal{M}}}(\beta) = \lim_{k\to\infty} \nu_k(\beta; b_{\mathcal{M}})$ are equal to $|d_\alpha|^2, \alpha = 0, 1$. Here, as usual, $\nu_k(\beta; b_{\mathcal{M}}) = n_k(\beta; b_{\mathcal{M}})/k$ is the relative frequency of realizations of the value $B = \beta$.

Remark 3.1. As we have already discussed, infinite statistical ensembles could not arise in any real physical experiment. We always operate with finite statistical ensembles (samples of finite lengths) S_N which are prepared by some preparation procedure $\mathcal{E}$ after N steps. However, a quantum state $\phi = \phi_{\mathcal{E}}$ (corresponding to this preparation procedure) cannot be considered as a representation of any of these finite ensembles S_N. A measurement for elements of S_N gives only a relative frequency, but not a probability. These frequencies may fluctuate when N is changed. Only asymptotically frequencies in S_N approach probabilities in S. Different properties may have different behaviour of fluctuations of frequencies before stabilization[11].

In the same way quantum states $\phi_\alpha = u_{\alpha 0}\psi_0 + u_{\alpha 1}\psi_1, \alpha = 0, 1$ describe some ideal infinite statistical ensembles $\bar{S}(A = \alpha)$ of quantum systems. In particular, these ensembles have the following frequency properties:

[11] I do not agree with the viewpoint of A. N. Kolmogorov: "The frequency concept based on the notion of *limiting frequency* as the number of trials increases to infinity does not contribute anything to substantiate the application of the results of probability theory to real practical problems where we always have to deal with a finite number of trials."

(1α) Any series of measurements $\mathcal{N}$ of the property A for elements $s_j \in \bar{S}(A = \alpha), j = 1, 2, ...,$ induces a collective

$$w_{\alpha\mathcal{N}} = (\theta_1, \theta_2, ..., \theta_k, ...),\ \theta_j = 0, 1,$$

such that frequency probabilities $\mathbf{P}_{w_{\alpha\mathcal{N}}}(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; w_{\alpha\mathcal{N}}) = 1$ and $\mathbf{P}_{w_{\alpha\mathcal{N}}}(1-\alpha) = \lim_{k\to\infty} \nu_k(1-\alpha; w_{\alpha\mathcal{N}}) = 0, \alpha = 0, 1.$

(2α) Any series of measurements $\mathcal{M}$ of the property B for elements $s_j \in \bar{S}(A = \alpha), j = 1, 2, ...,$ induces a collective

$$b_{\alpha\mathcal{M}} = (\lambda_1, \lambda_2, ..., \lambda_k, ...),\ \lambda_j = 0, 1, \tag{3.4}$$

such that frequency probabilities $\mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta) = \lim_{k\to\infty} \nu_k(\beta; b_{\alpha\mathcal{M}})$ are equal to $|u_{\alpha\beta}|^2,\ \beta = 0, 1.$

As we have already seen, by quantum calculus $\mathbf{P}_b(\beta) = |d_\beta|^2 = |c_0 u_{0\beta} + c_1 u_{1\beta}|^2, \beta = 0, 1.$ As in the case of ensemble probabilities, if we forget about dependence of probabilities on collectives and identify in the formula of total probability conditional probabilities $\mathbf{P}(\beta/\alpha) \equiv \mathbf{P}(B = \beta/A = \alpha), \alpha, \beta = 0, 1,$ with probabilities $\mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta)$ (of the property $B = \beta$ in the collectives $b_{\alpha\mathcal{M}}, \alpha = 0, 1$), then we arrive to the contradiction (between classical and quantum probability calculi). However, in the frequency approach there are even less reasons for such identification of probabilities than in the ensemble approach. Moreover, there immediately arises the problem of the correct frequency definition of conditional probabilities $\mathbf{P}(\beta/\alpha)$.

These conditional probabilities can be defined on the basis of the operation of combining of collectives, see section 9, Chapter 1. However, it is not clear which collectives

$$\bar{a} = (\alpha_1, \alpha_2, ..., \alpha_n, ...),\ A = \alpha_j = 0, 1,$$

and

$$\bar{b} = (\beta_1, \beta_2, ..., \beta_n, ...),\ B = \beta_j = 0, 1,$$

we have to combine to obtain conditional probabilities $\mathbf{P}(\beta/\alpha)$ (which must be equal to probabilities $\mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta)$ for observed collectives (3.4)). In any case the direct combination of observed collectives $a_{\mathcal{N}}$ and $b_{\mathcal{M}}$ would not produce such conditional probabilities, because these collectives are independent and here $\mathbf{P}(\beta/\alpha) = \mathbf{P}_{b_{\mathcal{M}}}(\beta) \neq \mathbf{P}_{b_{\alpha\mathcal{M}}}(\beta)$.

Moreover, we have shown (see section 9, Chapter 1), that (in the case of strictly positive probabilities) the condition of combining of $\bar{a}$ and $\bar{b}$ is

equivalent to the existence of a two dimensional collective

$$z = (z_1, z_2, ..., z_n, ...), \; z_j = (\alpha_j, \beta_j),$$

where α_j and β_j are elements of $\bar{a}$ and $\bar{b}$, respectively. Hence the two dimensional probability distribution

$$\mathbf{P}_z(\alpha, \beta) \equiv \mathbf{P}_z(A = \alpha, B = \beta) = \lim_{k \to \infty} \nu_k((\alpha, \beta); z)$$

must be well defined. Here $\nu_k((\alpha, \beta); z) = n_k((\alpha, \beta); z)/k, \; \alpha, \beta = 0, 1$, are relative frequencies of the realization of two dimensional labels $\gamma = (\alpha, \beta)$ in the collective z. We recall that the two dimensional probability distribution $\mathbf{P}_z(\alpha, \beta)$ and conditional probabilities are connected as $\mathbf{P}_z(\alpha, \beta) = \mathbf{P}_z(\beta/\alpha)\mathbf{P}_{\bar{a}}(\alpha)$. Thus if we assume that probabilities $\mathbf{P}_{\bar{a}}(\alpha)$ and conditional probabilities $\mathbf{P}_z(\beta/\alpha)$ are obtained on the basis of observed quantities, then we must assume that the two dimensional probability distribution $\mathbf{P}_z(\alpha, \beta)$ can be also obtained on the basis of observed quantities. Hence we have to assume the possibility of the simultaneous measurement of the pair $(A = \alpha, B = \beta)$. However, if the properties A and B are incompatible (represented by noncommuting operators $\hat{A}, \hat{B}$), then the existence of the simultaneous distribution for $(A = \alpha, B = \beta)$ contradicts to the quantum formalism.

Conclusion. *In the frequency probability framework it seems to be impossible to define conditional probabilities $\mathbf{P}_z(B = \beta/A = \alpha)$ on the basis of combining of collectives generated by some observations of incompatible properties A and B. Therefore the formula of total probability cannot be used for frequency physical probabilities.*

4. Kolmogorov formalism and quantum measurements. We consider now Kolmogorov's (ensemble-frequency) interpretation of 'quantum probabilities'. If we use the abstract measure-theoretical formalism, then we might identify some probabilities related to different probability spaces. This imply the contradiction between 'classical' and 'quantum' probabilities. In fact, different preparation procedures $\mathcal{G}$ are described by different probability spaces $(\Omega_{\mathcal{G}}, \mathcal{F}_{\mathcal{G}}, \mathbf{P}_{\mathcal{G}})$.

The quantum state $\phi = c_0\phi_0 + c_1\phi_1$, see (1.1), which describes the ensemble S (consisting of a statistical mixture of quantum systems with property $A = \alpha = 0, 1$ with probabilities $\mathbf{P}_\phi(A = \alpha) = |c_\alpha|^2$) is prepared via a preparation procedure $\mathcal{E}$. It is described by a Kolmogorov probability space $\mathcal{P} = (\Omega_{\mathcal{E}}, \mathcal{F}_{\mathcal{E}}, \mathbf{P}_{\mathcal{E}})$. The states $\phi_\alpha, \alpha = 0, 1$, which describe ensembles $\bar{S}(A = \alpha)$ (consisting of quantum systems with the definite value α of the

property A) are prepared via other preparation procedures $\mathcal{E}_\alpha$ (filters with respect to $A = \alpha$). These states must be described by other Kolmogorov probability spaces $(\Omega_{\mathcal{E}_\alpha}, \mathcal{F}_{\mathcal{E}_\alpha}, \mathbf{P}_{\mathcal{E}_\alpha})$.

Typically physicists apply the formula of the total probability by mixing conditional probabilities $\mathbf{P}_\mathcal{E}(\beta/\alpha)$ with respect to the probability space $\mathcal{P}$ with probabilities $\mathbf{P}_{\mathcal{E}_\alpha}(\beta), \alpha = 0, 1$. Such a manipulation induces the contradiction between 'classical' and 'quantum' probabilities.

5. Interference. Here we do not present statistical models which might explain the interference phenomena on the basis of the corpuscular picture (see [53], [54], [71]). We want just to illustrate our analysis of the notion of probability in quantum formalism by the example of the two slit experiment. This is the simplest experiment for demonstrating interference of light. There is a point source of light O and two screens L and L'. The screen L contains two slits h_0 and h_1. Light passes through S (through slits) and finally reaches the screen L' where the interference rings are observed. The wave explanation of the existence of interference rings is well known: the light reaching L' can travel by one of two routes–either through h_0 or through h_1; but the distances travelled by lights waves following these two paths are not equal and the light waves do not generally arrive at the screen 'in step' with each other.

On the other hand, O is a source of quantum particles, photons. To exclude the interaction between photons in a beam, we perform the experiment with very weak light, so that at any time there is only one photon in the region between O and L. The screen L' is replaced by a photographic plate or film (also denoted by L'). Individual spots appear on L' more or less chaotically. However, there appear standard interference fringes for a sufficiently long exposure. By this experiment we can compute the probability distribution of spots on L'. Here the property $A = 0, 1$, is given by a slit which is passed by a photon. The property B is obtained by the discretization of a measurement of the position on the screen L'. Let D be a domain on the plane L'. We set $B = 1$ if a photon is observed in D and $B = 0$ if a photon is observed in $L' \setminus D$. The formal application of the formula of total probability gives that $\mathbf{P}(B = \beta) = \mathbf{P}(A = 0)\mathbf{P}(B = \beta/A = \alpha) + \mathbf{P}(A = 1)\mathbf{P}(B = \beta/A = 1)$. However, experimental data demonstrates the violation of this equality. Mainly physicists interpret this violation as 'nonclassical behaviour' of photons. They claim that a photon does not pass one fixed slit.

Our ensemble analysis of the quantum formalism implies that we have to consider three ensembles: (1) S consists of all particles that pass through the

screen L when both slits are open; (2) $\bar{S}_0$ consists of all particles that pass through the screen L when the slit h_0 is open and the slit h_1 is closed; (3) $\bar{S}_1$ consists of all particles that pass through the screen L when the slit h_1 is open and the slit h_0 is closed. These ensembles of particles are represented by quantum states ϕ and ϕ_0, ϕ_1, respectively. In fact, probabilities which physicists use in the formula of the total probability for the two slit experiment are related to different ensembles of particles: $\mathbf{P}(A = \alpha) = \mathbf{P}_S(A = \alpha)$ and $\mathbf{P}(B = \beta/A = \alpha) = \mathbf{P}_{\bar{S}_\alpha}(B = \beta)$. Of course, the formula of total probability must hold true if instead of probabilities $\mathbf{P}_{\bar{S}_\alpha}(B = \beta)$ we would use probabilities $\mathbf{P}_S(B = \beta/A = \alpha)$. However, to find latter probabilities, we have to use sub-ensembles $S_\alpha, \alpha = 0, 1$, of the ensemble S consisting of particles that pass slits h_α. These sub-ensembles could not be found without to disturb the property B (because to find a slit, we have to perform the additional measurement, which, of course, change the distribution of B). Therefore it is insensible to discuss experimental verification of the formula of total probability for the two slit experiment.

In the following frequency analysis we shall use the framework of preparation/measurement. We consider preparation procedures $\mathcal{E}, \mathcal{E}_0, \mathcal{E}_1$ corresponding to the following configurations of open slits: h_0 and h_1 are open, h_0 is open and h_1 is closed, h_1 is open and h_0 is closed. The measurement procedure $\mathcal{M}$ is a measurement of the position B on the screen L'. Our frequency analysis of the quantum formalism implies that we have to consider three different collectives $b_{\mathcal{M}}$, $b_{0\mathcal{M}}$, $b_{1\mathcal{M}}$ which are obtained by the measurement $\mathcal{M}$ for particles prepared by $\mathcal{E}, \mathcal{E}_0, \mathcal{E}_1$, respectively. There are no reasons to identify probabilities $\mathbf{P}(B = \beta/A = \alpha)$ with frequency probabilities in the collectives $b_{\alpha\mathcal{M}}, \alpha = 0, 1$ (but typically they are identified). Moreover, probabilities $\mathbf{P}(B = \beta/A = \alpha)$ may be not well defined in the standard frequency framework.

If we keep to i-realism, then our probability analysis of the two slit experiment implies a *nonlocal* dependence on an equipment which is used for a preparation procedure (if we do not want to accept the special 'quantum probability rules'). The difference between probability distributions of B in the ensembles S_α and $\bar{S}_\alpha$ implies that the local change of the experimental arrangement (open slit or closed slit) implies global consequences. For example, if we close the slit h_0, then behaviour of the interaction between a photon and the slit h_1 is also changed. Thus, in fact, a photon interacts with the whole screen L. The same consequence we obtain in the frequency framework.

However, if we keep to f-realism, then a problem of non-locality does not arise.

Here we do not assume that initial values A_i of the property A are observed. Probabilities given by quantum states ϕ_α (by ensembles $\bar{S}_\alpha$) are conditional probabilities with respect to final values A_f of A, $\mathbf{P}(B = \beta / A_f = \alpha)$. But in the formula of the total probability for the quantum state ϕ (the ensemble S) we have to use probabilities $\mathbf{P}(A_i = \alpha)$ and $\mathbf{P}(B = \beta / A_i = \alpha)$. The values A_i for the quantum state ϕ (the ensemble S) and the corresponding probabilities cannot be found in the experimental framework. Analysis of such probability distributions with respect to a 'hidden property' will be presented in section 4.

6. Non-ergodic interpretation of quantum mechanics. We discuss now other delicate problem in the probabilistic foundations of quantum mechanics. As it has been pointed out, Kolmogorov's (ensemble-frequency) interpretation of probability implies identification of the ensemble and frequency approaches to probability. As a consequence, it is always assumed that frequency probabilities in the collective $b_{\mathcal{M}}$, see (3.3), can be identified with probabilities with respect to the ensemble $\mathcal{S}$ of pairs (a, u), where a is a quantum particle and u is an equipment which is used to measure the property B of a.[12] However, this postulate has never been tested experimentally[13]. Brasilian physicist N. Buonumano proposed a *non-ergodic interpretation* of quantum mechanics, [17]. By this interpretation frequency probabilities need not coincide with $\mathcal{S}$-ensemble probabilities. This imply that in principle trials need not be independent (see [53], [54]). Thus there may be correlations between x_i and $x_j, i \neq j$, in x (of course, in this case x would not be Mises' collective).

4 Probabilities with respect to objective conditions

The formalism of quantum mechanics implies that it is impossible to perform experiments for a simultaneous measurement $(A = \alpha, B = \beta)$ of two incompatible properties A and B of a quantum system[14]. Therefore two di-

[12]Despite of the fact that in all real experiments this collective is generated by the long chain of successive experiments with the same equipment u_{fix}.

[13]The group of H. Rauch at Atominstitute in Wien did some indirect experiments in this direction, see [96] (the most interesting experiment was performed by J. Summhammer, [101]).

[14]In fact, on the physical level incompatible properties are defined as properties for which it is impossible to perform a simultaneous measurements. The representation of such properties by noncommuting operators in a Hilbert space of quantum states is only

mensional probabilities $\mathbf{P}(A = \alpha, B = \beta)$ (or equivalently conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ for $\mathbf{P}(A = \alpha) > 0$) could not be found on the experimental basis.

However, it is still sensible to discuss the existence of such probabilities if it is supposed that properties A, B are objective. Of course, the simultaneous existence of two objective properties does not imply automatically the possibility to perform a simultaneous measurement of these properties.

We shall study the most general situation. It is supposed that B is some observed property (it may be objective or created by the act of a measurement) and A is some objective property which cannot be observed simultaneously with B. We shall study the problem of existence of conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ in the frequency and ensemble frameworks. The present scheme covers different approaches to properties of quantum systems:

(i) We may keep to i-realism. Here the observed value of B coincides with its initial value. Thus we study probabilities $\mathbf{P}(B_i = \beta/A_i = \alpha)$ where A_i and B_i are initial values of properties.

(f) We may keep to f-realism. Here the observed value B_f of the property B can differ from its initial value B_i. Thus we study probabilities $\mathbf{P}(B_f = \beta/A_i = \alpha)$.

(e) We may keep to empiricism (or even idealism). Here the property B is created by the act of a measurement. It is meaningless to speak about values of B before a measurement.

In fact, f-realism seems to be the most attractive. Here we can use the preparation/measurement approach. Some preparation procedure $\mathcal{E}$ produces a statistical ensemble S of particles with the definite probability distribution for the property A (which is typically supposed to be objective). Then a measurement $\mathcal{M}$ of other property B is performed for particles $s \in S$. This measurement gives final values B_f of the property B.

The main result of our considerations is that conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$ and two dimensional probabilities $\mathbf{P}(A = \alpha, B = \beta)$ may be not exist (both in frequency and ensemble approaches). This result seems to be rather strange from the Kolmogorov probability viewpoint.

Our models in that $\mathbf{P}(B = \beta/A = \alpha)$ do not exist give examples of (non-Kolmogorovean) probabilistic models without conditional probabilities (probabilities are well defined, but conditional probabilities could not be

a consequence of such impossibility of simultaneous measurements.

defined). Here probability is not closed. It is defined on the system of events which do not form a set algebra.

1. Frequency probabilities. To define frequency conditional probabilities $\mathbf{P}(B = \beta/A = \alpha)$, we must combine two collectives a and b corresponding to values of A and B. We choose a collective $b = b_{\mathcal{M}}$, see (3.3), induced by a measurement $\mathcal{M}$ of the property B as one of collectives for combining. As the property A is objective, then each element $s_j \in S$ has this property. Hence parallel to the constructing of the collective $b_{\mathcal{M}}$ we can imagine the process of construction of a 'hidden sequence'

$$a_h = (\alpha_1, \alpha_2, ..., \alpha_n, ...), \ \alpha_j = 0, 1,$$

where $\alpha_j = \alpha$ if the property A has the value α for a quantum system s_j. Suppose that the sequence a_h is a collective. We choose $a = a_h$ as another collective for combining. Suppose that probability distributions $\mathbf{P}_b(\beta)$ and $\mathbf{P}_a(\alpha)$ are strictly positive. Finally we suppose that collectives a and b are combinable. Therefore the two dimensional sequence

$$z = (z_1, z_2, ..., z_j, ...), \ z_j = (\alpha_j, \beta_j),$$

corresponding to these collectives is also a collective. Hence frequency conditional probabilities $\mathbf{P}_z(\beta/\alpha) \equiv \mathbf{P}_z(B = \beta/A = \alpha)$ are defined via the standard scheme:

Suppose that there are $M_k(\alpha; z)$ elements with the first coordinate α among the first k elements of z, and there are $n_k(\beta/\alpha; z)$ elements with the first coordinate β among these $M_k(\alpha; z)$ elements. We introduce the relative frequencies:

$$\nu_k(\alpha; z) = \frac{M_k(\alpha; z)}{k} \ \text{ and } \ \nu_k(\beta/\alpha; z) = \frac{n_k(\beta/\alpha; z)}{M_k(\alpha; z)} \ .$$

Conditional probability is defined as

$$\mathbf{P}_z(\beta/\alpha) = \lim_{k \to \infty} \nu_k(\beta/\alpha; z).$$

This definition can be reformulated in the following way. For each fixed $\alpha = 0, 1$, we choose a subsequence

$$b_\alpha = (\beta_1, \beta_2, ..., \beta_n, ...), \ \beta_j = 0, 1,$$

of the sequence z consisting of second coordinates of $z_j = (\alpha_j, \beta_j)$ with $\alpha_j = \alpha$. Then $\mathbf{P}_z(\beta/\alpha) = \lim_{k \to \infty} \nu_k(\beta; b_\alpha)$.

The conditional probability $\mathbf{P}_z(\beta/\alpha)$ has the following meaning: it is probability to observe the value β of the property B under the condition that the hidden (but objectively existing) value of the property A is equal to α. The two dimensional probability distribution $\mathbf{P}_z(\alpha,\beta) = \mathbf{P}_a(\alpha)\mathbf{P}_z(\beta/\alpha)$ is also well defined. This probability has the following physical meaning: it is probability that the hidden property $A = \alpha$ and the observed property $B = \beta$.

2. No conditional probabilities, no Bayes' formula. In principle there may be some frequency 'pathologies'. It may be that the hidden sequence $a = a_h$ is not a collective or it is a collective, but the collectives a and b are not combinable (physical experience implies that the sequence $b = b_{\mathcal{M}}$ is always a collective). Let us analyse such a situation more carefully. To simplify our considerations, we start with the case in that $a = a_h$ is a collective. Here frequency probabilities $\mathbf{P}_a(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a)$ are well defined. However, we do not more assume that the collectives a and b are combinable.

We have $n_k(\beta; b) = n_k(\beta/0; z) + n_k(\beta/1; z)$. Thus we obtain

$$\nu_k(\beta; b) = \frac{n_k(\beta/0;z)}{M_k(0;z)}\frac{M_k(0;z)}{k} + \frac{n_k(\beta/1;z)}{M_k(1;z)}\frac{M_k(1;z)}{k} .$$

Thus we have

$$\nu_k(\beta; b) = \nu_k(\beta/0; z)\nu_k(0; a) + \nu_k(\beta/1; z)\nu_k(1; a) . \tag{4.1}$$

We also note that by our assumptions there exist $\mathbf{P}_b(\beta) = \lim_{k\to\infty} \nu_k(\beta; b)$ and $\mathbf{P}_a(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a)$. We ask the following question:

Is it possible that (despite of the existence of the above limits and despite of equality (4.1)) $\lim_{k\to\infty} \nu_k(\beta/\alpha; z)$ *do not exist?*

Yes, it is surely possible!

Example 4.1. Let $\mathbf{P}_a(\alpha) = \lim_{k\to\infty} \nu_k(\alpha; a) = 1/2, \alpha = 0, 1$. As always $\nu_k(0/\alpha; z) + \nu_k(1/\alpha; z) = 1$ for $\alpha = 0, 1$, it is possible to represent conditional frequencies in the form

$$\nu_k(0/\alpha; z) = \sin^2 \phi_{\alpha,k}, \ \nu_k(1/\alpha; z) = \cos^2 \phi_{\alpha,k} ,$$

where the *phase* $\phi_{\alpha,k} = \arcsin \sqrt{\nu_k(0/\alpha; z)}$. In the case of regular conditional behaviour angles $\phi_{\alpha,k}$ stabilize (mod 2π) to some values ϕ_α when $k \to \infty$. Here conditional probabilities are well defined: $\mathbf{P}_z(0/\alpha) = \sin^2 \phi_\alpha$

and $\mathbf{P}_z(1/\alpha) = \cos^2\phi_\alpha$. Equality (4.1) implies the formula of total probability:

$$\mathbf{P}_b(0) = \frac{1}{2}(\sin^2\phi_0 + \sin^2\phi_1), \;\; \mathbf{P}_b(1) = \frac{1}{2}(\cos^2\phi_0 + \cos^2\phi_1)\;.$$

Let us consider now the case of irregular conditional behaviour. Here angles $\phi_{\alpha,k}$ do not stabilize (mod 2π) when $k \to \infty$. But by (4.1) we have that limits

$$\mathbf{P}_b(0) = \frac{1}{2}\lim_{k\to\infty}(\sin^2\phi_{0,k} + \sin^2\phi_{1,k}), \;\; \mathbf{P}_b(1) = \frac{1}{2}\lim_{k\to\infty}(\cos^2\phi_{0,k} + \cos^2\phi_{1,k})$$

must exist. For example, these conditions can be satisfied if we choose $\phi_{1,k} \approx \frac{\pi}{2} - \phi_{0,k}, k \to \infty$. Thus there is no contradiction between nonexistence of frequency conditional probabilities and formula (4.1).

What is a physical meaning of fluctuations of conditional relative frequencies $\nu_k(\beta/\alpha)$ (nonexistence of conditional probabilities $\mathbf{P}_z(\beta/\alpha)$)?

A quantum state ϕ contains only information on asymptotic behaviour of frequencies for observations of each fixed property. However, ϕ does not contain information on statistical relations between different properties. This relation is given by conditional frequencies which are not determined by the quantum formalism. Therefore in principle behaviour of relative frequencies in the statistical ensemble S (represented by ϕ) may be extremely irregular. But these fluctuations of conditional frequencies may compensate one another and give well defined frequency probabilities for observed properties.

A real preparation procedure $\mathcal{E}$ can produce (after N steps) only a finite approximation S_N of the (ideal infinite) ensemble S represented by ϕ. Fluctuations of conditional frequencies imply that the statistical relation between two properties A and B (or more precisely the reaction of a quantum system s with the fixed (hidden) value α of the property A to a measurement of the property B) may strongly depend on the number N of experiments N. Let us consider again Example 4.1 and let $\phi_{0,k} \approx \frac{\pi k}{2m}, \phi_{1,k} \approx \frac{\pi}{2} - \frac{\pi k}{2m}, k \to \infty$, where $m > 1$ is the fixed natural number. Here 'conditional probabilities'

$$\mathbf{P}^k(0/0) \equiv \nu_k(0/0;z) \approx \sin^2\frac{\pi k}{2m}, \;\; \mathbf{P}^k(1/0) \equiv \nu_k(1/0;z) \approx \cos^2\frac{\pi k}{2m},$$

$$\mathbf{P}^k(0/1) \equiv \nu_k(0/1;z) \approx \cos^2\frac{\pi k}{2m}, \;\; \mathbf{P}^k(1/1) \equiv \nu_k(1/1;z) \approx \sin^2\frac{\pi k}{2m}$$

oscillate with the period $T = 2m$, when $k \to \infty$. Let m be very large. Then, for $k = 2mj + 1, \mathbf{P}^k(0/0) = \mathbf{P}^k(1/1) \approx 0$ and $\mathbf{P}^k(1/0) = \mathbf{P}^k(0/1) \approx 1$.

Therefore in the ensemble S_k practically every quantum system s having the property $A = 0$ will exhibit the property $B = 1$ and practically every quantum system s having the property $A = 1$ will exhibit the property $B = 0$ (in the measurement $\mathcal{M}$ of B.) However, after $(m - 1)$ steps statistical conditional behaviour changes crucially. For $k' = 2mj+m, \mathbf{P}^k(0/0) = \mathbf{P}^k(1/1) \approx 1$ and $\mathbf{P}^k(1/0) = \mathbf{P}^k(0/1) \approx 0$. Therefore in the ensemble $S_{k'}$ practically every quantum system s having the property $A = 0$ will exhibit the property $B = 0$ and practically every quantum system s having the property $A = 1$ will exhibit the property $B = 1$. At the same time observed 'probabilities' $\mathbf{P}^k(B = \beta) = \nu_k(\beta; b)$ do not depend on these oscillations of conditional probabilities.

Remark 4.1. (On fluctuations of ensemble conditional probabilities) The above arguments can be used for ensemble conditional probabilities. A quantum state ϕ represents an *infinite ideal ensemble* S of quantum systems. As we have already discussed in section 3, real analysis does not give a possibility to use the proportional definition of probability with respect to S. Typically probabilities $\mathbf{P}_S(B = \beta)$ with respect to S are considered as limits of probabilities $\mathbf{P}_{S_N}(B = \beta)$ with respect to finite approximations S_N of S. Such an approach to probabilities $\mathbf{P}_S(B = \beta)$ is justified by the incredible number of quantum experiments. However, it is often supposed that conditional probabilities $\mathbf{P}_S(B = \beta/A = \alpha)$ can be also defined as limits of probabilities $\mathbf{P}_{S_N}(B = \beta/A = \alpha)$ with respect to finite approximations S_N of S. Such an assumption has not been (and probably it will never be) verified experimentally. Example 4.1 (which can be used in the ensemble framework) demonstrates that in principle conditional probabilities $\mathbf{P}_{S_N}(B = \beta/A = \alpha)$ may oscillate with the increasing of N. In such a case conditional probabilities $\mathbf{P}_S(B = \beta/A = \alpha)$ cannot be defined. Therefore Bayes' formula and the formula of total probability cannot be used for such quantum states.

Finally we remark that it may be that the sequence $a = a_h$ is not collective. For example, if we keep to f-realism, then the statistical stabilization of frequencies $\nu_k(A_f = \alpha)$ need not imply the statistical stabilization of frequencies $\nu_k(A_i = \alpha)$.

Example 4.2. (Fluctuating probabilities and stabilized conditional probabilities). Suppose that $\nu_k(0, a) \approx \sin^2 \phi_k$ and $\nu_k(1, a) \approx \cos^2 \phi_k, k \to \infty$. If phases ϕ_k do not stabilize (mod 2π) when $k \to \infty$, then frequencies $\nu_k(\alpha; a)$ fluctuate when $k \to \infty$. Thus frequency probabilities $\mathbf{P}_a(\alpha)$ do not exist. Suppose that, however, frequency conditional probabilities $\mathbf{P}_z(\beta/\alpha) = \lim_{k\to\infty} \nu_k(\beta/\alpha; z)$ exist and they are equal to $1/2$. Therefore sizes of popu-

lations $S_{N,\alpha}$ with the fixed value $A = \alpha$ fluctuate, but reactions of quantum systems $s \in S_{N,\alpha}$ to the measurement $\mathcal{M}$ of the B are stable. In this case we find that the limit in (4.1) exists

$$\mathbf{P}_b(\beta) = \lim_{k \to \infty} (\nu_k(0; a)\nu_k(\beta/0; z) + \nu_k(1; a)\nu_k(\beta/1; z)) = 1/2, \ \beta = 0, 1.$$

Thus there is no contradiction between nonexistence of frequency probabilities $\mathbf{P}_a(\alpha)$ and formula (4.1).

5 Einstein-Podolsky-Rosen paradox: probability, reality and locality

In the previous sections we have shown that there is no principal difference between 'quantum' and 'classical' probabilities and as consequence between classical and quantum systems. We can use ensemble or frequency definitions of probability. However, we have to control the relation between probabilities and ensembles or collectives. On the other hand, we cannot use the conventional Kolmogorov formalism in that the structure of an ensemble or a collective does not play any role. In principle, it is possible to consider 'quantum properties' as objective properties (by using both i-realism and f-realism). Of course, probabilistic distributions of these properties depend on ensembles or collectives.

However, there are some quantum experiments which seem to demonstrate that there is a large difference between classical and quantum systems. All such experiments are based on the idea to eliminate disturbance effects by separating quantum systems in space-time (and to use correlations between these separated quantum systems). The starting point was the famous Einstein-Podolsky-Rosen (EPR) experiment [36]. We present a brief description of this experiment. We start with the definition of **einsteinian separability:**

Two space-time regions U and V are said to be spatially separated, if the real factual situation within V is independent of what is done in U.

We also remark that A. Einstein was (more or less) an adherent of i-realism. Thus values of physical properties which will be discussed later (namely, positions $\mathbf{q}$ and momentums $\mathbf{p}$) are initial values of this properties.

It should be noticed that the study of distinguishing features of 'quantum probabilities' was not the original aim of EPR's considerations. EPR wanted

to show that quantum mechanics is not a *complete physical theory.* The completeness means that quantum mechanics provides a complete description of the atomic and subatomic phenomena. The opinion that quantum mechanics is complete (and hence we need not more detailed description of reality than quantum mechanics) was, already at that time (1933), so much engrained in the mind of physicists that the EPR arguments against the completeness was soon referred as a paradox. EPR wanted to demonstrate that there exist elements of reality which could not be described by a quantum state. Of course, in this framework the question on the meaning of an **element of reality** arose immediately. EPR thought that it would be impossible to propose the exact definition of an element of reality. However, they proposed the following criterion of reality:

"If, without in any way disturbing a system we can predict with certainty (i.e., with probability equal to unity) the value of a physical quantity, then there exists an element of physical reality corresponding to this physical quantity."

EPR proposed the following arguments based on this criterion for elements of reality and the notion of separability for two space-time regions U and V.

Let us consider a statistical ensemble S of pairs (a^1, a^2) of correlated particles. For example, these are pairs of particles which are emitted by excited atoms. We consider the one dimensional model with particles moving in the opposite directions. For each pair the correlation implies the conservation of the momentum, $\mathbf{p}^{(1)} + \mathbf{p}^{(2)} = 0$, and the relative position, $\mathbf{q}^{(1)} - \mathbf{q}^{(2)} = 0$ (correlations between properties of particles). For any pair (a^1, a^2) of correlated particles, we can measure the position $\mathbf{q}^{(1)}$ of a^1 in U which (due to the correlation) gives the position $\mathbf{q}^{(2)}$ of a^2 (without to disturb a^2). Thus the position $\mathbf{q}^{(2)}$ of a^2 is an element of reality. By the similar considerations we obtain that the momentum $\mathbf{p}^{(2)}$ of a^2 is an element of reality. On the other hand, the quantum formalism implies the Heisenberg uncertainty relation

$$\Delta\mathbf{q}\Delta\mathbf{p} \geq h/2$$

for any quantum state ϕ. Thus the definite values of the position and momentum of a quantum particle cannot be simultaneously elements of reality for the same quantum state. EPR interpreted this as the evidence of the incompleteness of quantum mechanics: in the EPR experiment two elements of reality (the position and momentum) exist simultaneously, but they could

not be described by any quantum state (thus the formalism of quantum mechanics does not provide the description of the whole physical reality).

The EPR considerations (which are often regarded as the paradox in the foundations of quantum mechanics) induced great debates (which were initiated by A. Einstein and N. Bohr). Numerous arguments were used by both sides. It is interesting to remark that at the first stage of these debates probability reasons were not used. There was no analysis of the probability basis of the EPR considerations. In particular, nobody tried to study the problem of difference between 'classical and quantum probabilities' to disprove the simultaneous reality of the position and momentum. However, later such analysis has been done and one of the results of this analysis was famous Bell's inequality (see section 6).

I support the viewpoint that quantum mechanics is not complete. The incompleteness of quantum mechanics is rather a consequence of all physical experience which demonstrated that no physical theory (at least before quantum mechanics) turned out to be universally valid. Every single theory was valid only if applied to a restricted part of reality, its domain of application. Do we have any reason to believe that quantum mechanics is different, and will hold true for whatever future experiments we may be able to think of? But at the same time I think that EPR arguments do not imply the conclusion that quantum mechanics is not complete. I am not satisfied by EPR's criterion of reality. There are strong probabilistic arguments against this criterion. The meaning of 'unit probability' in this criterion is unclear. In fact, this 'unit probability' must depend on an ensemble or a collective. We shall see that it is impossible to find the same ensemble or collective for the positions and momentums of particles a^2 (or particles a^1).

To save completeness of quantum mechanics, some physicists accept *nonlocality of space-time.* They claim that, for example, a measurement of the position of the particle a^1 located in Moscow changes properties of the particle a^2 located in Vladivostok. Some of them assume the possibility of the *faster-than-light-influences* (of course, such an assumption contradicts to theory of relativity). Other adherents of nonlocality consider this nonlocality as only *information nonlocality.* They think that, for example, a measurement of the position $\mathbf{q}^{(1)}$ of the particle a^1 does not change objective properties of the particle a^2, but such a measurement changes our information about the particle a^2. Hence they need not use the *faster-than-light-influences.*

Another group of physicists thinks that the root of the problem is the realist viewpoint of EPR. If we reject realism and keep to empiricism (or

even idealism), then we could not assign any physical meaning to values of the position $\mathbf{q}^{(2)}$ and momentum $\mathbf{p}^{(2)}$ of the particle a^2 which are predicted on the basis of measurements for the particle a^1. Adherents of empiricism have also some differences in views. One part of them think that the root of the problem is the impossibility to perform a simultaneous measurement of the position $\mathbf{q}^{(1)}$ and momentum $\mathbf{p}^{(1)}$ for the same particle a^1 (and thus obtain the 'simultaneous prediction'). Other keep to the rigid empiricists line. They think that the root of the problem is the impossibility to perform a simultaneous measurement of the position $\mathbf{q}^{(2)}$ and momentum $\mathbf{p}^{(2)}$ for the same particle a^2.

As I have already mentioned, it seems that EPR arguments do not imply incompleteness, nonlocality or impossibility to keep to realism. EPR considerations imply only that we could not manipulate with formal (abstract) probabilities which are not related to concrete ensembles or collectives.

By fixing a value $\alpha \in \mathbf{R}$ of $\mathbf{q}^{(1)}$ we can construct an ensemble $\bar{S}_\alpha$ of particles a_j^2 for that $\mathbf{q}^{(2)} = \alpha$ and the distribution of the momentums is the same as in the original ensemble. Of course, probability $\mathbf{P}_{\bar{S}_\alpha}(\mathbf{q}^{(2)} = \alpha) = 1$. Thus $\mathbf{q}^{(2)} = \alpha$ is an element of reality for the ensemble $\bar{S}_\alpha$. However, $\mathbf{P}_{\bar{S}_\alpha}(\mathbf{p}^{(2)} = \beta) \neq 1$ for any fixed value $\beta \in \mathbf{R}$. Thus $\mathbf{p}^{(2)} = \beta$ is not an element of reality for this ensemble.

In the same way by fixing the value $\beta \in \mathbf{R}$ of $\mathbf{p}^{(1)}$ we can construct an ensemble $\bar{R}_\beta$ of particles a_j^2 for that $\mathbf{p}^{(2)} = \beta$ and the distribution of the positions is the same as in the original ensemble. Probability $\mathbf{P}_{\bar{R}_\beta}(\mathbf{p}^{(2)} = \beta) = 1$. Thus $\mathbf{p}^{(2)} = \beta$ is an element of reality for the ensemble $\bar{R}_\beta$. However, $\mathbf{P}_{\bar{R}_\beta}(\mathbf{q}^{(2)} = \alpha) \neq 1$ for any fixed value $\alpha \in \mathbf{R}$. Thus $\mathbf{q}^{(2)} = \alpha$ is not an element of reality for this ensemble. EPR did not present any idea how we could construct an ensemble $W_{\alpha\beta}$ of particles a_j^2 such that $\mathbf{P}_{W_{\alpha\beta}}(\mathbf{q}^{(2)} = \alpha) = 1$ and $\mathbf{P}_{W_{\alpha\beta}}(\mathbf{p}^{(2)} = \beta) = 1$. Therefore the EPR arguments give no reason to conclude that quantum mechanics is not complete. There are no reasons to use nonlocality or to reject realism to explain the EPR arguments. It must be pointed out that EPR arguments could not be used as a 'proof' that i-realism can (or even must) be used to describe quantum phenomena[15]. In fact, from the mathematical point of view the only 'argument' of EPR was

[15]EPR obtained incompleteness of quantum mechanics by presenting the experiment which demonstrates that both the position and momentum of a quantum particle can be elements of reality for the same quantum state. This is often interpreted as a proof of the possibility to use i-realists approach in quantum mechanics.

that the notion of the unit probability can be used without connection to concrete ensembles.

Our previous considerations can be repeated in the frequency framework. Here we can keep not only to i-realism, but also to f-realism in the EPR scheme[16]. Let us perform a measurement $\mathcal{N}$ of $\mathbf{q}^{(1)}$ and a measurement $\mathcal{M}$ of $\mathbf{p}^{(2)}$ and save only the results for that $\mathbf{q}^{(1)} = \alpha$, where $\alpha \in \mathbf{R}$, is some fixed value. We obtain the two dimensional collective: $x_\alpha = (x_1, x_2, ..., x_k, ...)$, $x_j = (\mathbf{q}_j^{(1)}, \mathbf{p}_j^{(2)})$, where $\mathbf{q}_j^{(1)} \equiv \alpha$. As $\mathbf{q}_j^{(2)} = \mathbf{q}_j^{(1)} \equiv \alpha$, we can (parallel to the construction of x_α) construct another two dimensional collective $x'_\alpha = (x'_1, x'_2, ..., x'_k, ...)$, $x'_j = (\mathbf{q}_j^{(2)}, \mathbf{p}_j^{(2)})$, where $\mathbf{q}_j^{(2)} \equiv \alpha$. In the same way, for each fixed value $\beta \in \mathbf{R}$ of the momentum, we construct two dimensional collectives $y_\alpha = (y_1, y_2, ..., y_k, ...)$, $y_j = (\mathbf{q}_j^{(2)}, \mathbf{p}_j^{(1)})$, where $\mathbf{p}_j^{(1)} \equiv \beta$, and $y'_\alpha = (y'_1, y'_2, ..., y'_k, ...)$, $y'_j = (\mathbf{q}_j^{(2)}, \mathbf{p}_j^{(2)})$, where $\mathbf{p}_j^{(2)} \equiv \beta$. Of course, $\mathbf{P}_{x'_\alpha}(\mathbf{q}^{(2)} = \alpha) = 1$ and $\mathbf{q}^{(2)} = \alpha$ is the element of reality for the collective x'_α and $\mathbf{P}_{y'_\beta}(\mathbf{p}^{(2)} = \beta) = 1$ and $\mathbf{p}^{(2)} = \beta$ is the element of reality for the collective y'_β. However, EPR did not present any idea how we can construct a collective $z_{\alpha\beta}$ such that $\mathbf{P}_{z_{\alpha\beta}}(\mathbf{q}^{(2)} = \alpha) = 1$ and $\mathbf{P}_{z_{\alpha\beta}}(\mathbf{p}^{(2)} = \beta) = 1$.

6 Bell's inequality for probabilities

The EPR idea to consider statistical ensembles of correlated spatially separated particles was developed by D. Bohm. He proposed a simpler example in that it is possible to use discrete variables.

1. EPR experiment for spin. Instead of the position and momentum of a quantum particle, he considered its spin[17] components. Let $\mathbf{s} \in \mathbf{R}^3$ be the spin of a quantum particle. For any axis $n \in \mathbf{R}^3$ we denote the projection of $\mathbf{s}$ to this axis by the symbol $\mathbf{s}_n$ (i.e., $\mathbf{s}_n = \frac{(\mathbf{s},n)n}{\|\mathbf{s}\|}$), where $(\cdot,\cdot)$ and $\|\cdot\|$ are, respectively, the inner product and norm on $\mathbf{R}^3$). There exists an equipment M_n for measuring the spin projection $\mathbf{s}_n$. However, such a measurement disturbs a quantum particle and changes the spin. There are no measurement devices $M_{n,n'}$ which can measure two components $\{\mathbf{s}_n, \mathbf{s}_{n'}\}$, $n \neq n'$, simultaneously. However, for correlated particles (a^1, a^2) (with spins $\mathbf{s}^1, \mathbf{s}^2$) we can

[16]In the latter case $\mathbf{q}^{(l)} \equiv \mathbf{q}_f^{(l)}$ and $\mathbf{p}^{(l)} \equiv \mathbf{p}_f^{(l)}, l = 1, 2$, and $\mathbf{q}_f^{(1)} - \mathbf{q}_f^{(2)} = 0$, $\mathbf{p}_f^{(1)} + \mathbf{p}_f^{(2)} = 0$.

[17]The scientists whose interests are far from quantum mechanics may imagine spin as an arrow $\mathbf{s} \in \mathbf{R}^3$ which is associated with each quantum particle indicating the 'internal rotation' of the particle.

use the conservation law for the spins of these particles: $s^1 + s^2 = 0$. Hence the measurement M_n^1 for a^1 gives automatically the value s_n^2 of the spin of a^2. As usual, it is assumed that particles a^1 and a^2 satisfy the condition of einsteinian separability. By the EPR reality criterion we obtain that there exists an element of reality corresponding to the spin component s_n^2 (and by symmetry for the s_n^1) for any axis $n \in \mathbf{R}^3$. Therefore the spin s is an element of reality.

By probabilistic reasons (discussed in the previous section) we do not apply the EPR reality criterion. However, we can study the following problem:

Is it possible to use realism for describing spin measurements for correlated particles?

We restrict our consideration to the plain model. Here each direction n can be characterized by an angle $\phi : n = n_\phi$. We set $s_\phi = \text{sign}(s, n_\phi)$. In the real physical model we have to use probabilities of simultaneous measurements of $s^1_{\phi_j}$ and $s^2_{\phi_l}$ for three angles $\phi_j, j = 1, 2, 3$. [18] It is possible to obtain some inequality for these probabilities, namely, *Bell's inequality.* As there are two particles a^1 and a^2, to describe the (plain) model we must use the four dimensional Hilbert space. However, we can obtain the same results on the basis of the two dimensional Hilbert space by using the following toy-model.

Let $\mathcal{H}$ be the two dimensional Hilbert space and let $e_\phi = (e_{\phi,+}, e_{\phi,-}), \phi \in [0, 2\pi)$, be orthonormal bases in $\mathcal{H}$ which are connected by the following unitary transformation:

$$e_{\phi,+} = \cos(\theta - \phi)e_{\theta,+} + i\sin(\theta - \phi)e_{\theta,-} \ , \tag{6.1}$$

$$e_{\phi,-} = i\sin(\theta - \phi)e_{\theta,+} + \cos(\theta - \phi)e_{\theta,-} \ . \tag{6.2}$$

We introduce the quantum state

$$\Psi = \frac{e^{i\gamma}}{\sqrt{2}}e_{\gamma,+} + \frac{e^{i\gamma}}{\sqrt{2}}e_{\gamma,-} \ .$$

We consider observables (properties) s_ϕ corresponding to bases e_ϕ : $\hat{s}_\phi e_\phi = \pm e_\phi$. By the probability interpretation of the quantum state Ψ we have $\mathbf{P}_\Psi(s_\phi = \pm 1) = 1/2$.

If (as it is usually assumed by physicists) conditional probabilities $\mathbf{P}(\omega \in \Omega : s_\theta(\omega) = \epsilon / s_\phi(\omega) = \delta), \epsilon, \delta = \pm 1$, are identified with probabilities given by

[18] In fact, in experiments we need to use even four angles.

expansions (6.1), (6.2), then we obtain the following 'quantum probabilities'

$$\mathbf{P}_q(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1, \mathbf{s}_\phi(\omega) = -1)$$

$$= \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1)\mathbf{P}(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1/\mathbf{s}_\phi(\omega) = -1) \tag{6.3}$$

$$= \tfrac{1}{2}\sin^2(\theta - \phi)$$

and

$$\mathbf{P}_q(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1, \mathbf{s}_\phi(\omega) = +1)$$

$$= \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = +1)\mathbf{P}(\omega \in \Omega : \mathbf{s}_\theta(\omega) = +1/\mathbf{s}_\phi(\omega) = +1) \tag{6.4}$$

$$= \tfrac{1}{2}\cos^2(\theta - \phi).$$

It is interesting to note that there is no contradiction between 'quantum and classical probability rules' in this model. The reader can easily check the validity of Bayes' formula.

2. Bell's inequality for probabilities. We prove now some inequality for events defined by three variables $\mathbf{s}_\gamma(\omega)$, $\gamma = 0, \phi, \theta$. In fact, this inequality does not depend on the form of the probability distributions of random variables $\mathbf{s}_\gamma(\omega)$. We shall use only the fact that there exists the Kolmogorov probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ on which these random variables are defined:

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1)$$

$$= \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \tag{6.5}$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = -1),$$

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1)$$

$$= \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \tag{6.6}$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = -1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1),$$

and

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1)$$

$$= \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \tag{6.7}$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1).$$

If we add together the equations (6.5) and (6.6) we obtain

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) + \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1)$$

$$= \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = +1)$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = -1)$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1)$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = -1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1). \tag{6.8}$$

But the first and the third terms on the right hand side of this equation are just those which when added together make up the term $\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1)$ (Kolmogorov probability is additive). It therefore follows that:

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) + \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1)$$

$$= \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1)$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1, \mathbf{s}_\theta(\omega) = -1)$$

$$+\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = -1, \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1) \tag{6.9}$$

By using nonnegativity of probability we obtain the inequality:

$$\mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\phi(\omega) = +1) + \mathbf{P}(\omega \in \Omega : \mathbf{s}_\phi(\omega) = -1, \mathbf{s}_\theta(\omega) = +1)$$

$$\geq \mathbf{P}(\omega \in \Omega : \mathbf{s}_0(\omega) = +1, \mathbf{s}_\theta(\omega) = +1) \tag{6.10}$$

which is a variant of Bell's inequality (for probabilities).

We turn back to physics and apply the inequality (6.10) to the 'quantum probabilities' $\mathbf{P}_q$, see (6.3), (6.4), which were computed in the framework of quantum mechanics. We obtain: $\cos^2\phi + \sin^2(\theta - \phi) \geq \cos^2\theta$. Now set $\phi = 3\theta$. We obtain: $g(\theta) = \cos^2 3\theta + \sin^2 2\theta - \cos^2\theta \geq 0$. However, the latter inequality holds only for *sufficiently large* angles $\theta : \theta \geq \pi/6$. Thus for $\theta < \pi/6$ the inequality (6.10) is violated.

7 Bell's mystification

First of all we remark that the violation of Bell's inequality for 'quantum probabilities' may be in principle interpreted as an evidence of violations of quantum mechanical laws for the spin model. However, numerous experiments were performed in the connection with this problem, see, for example, [5], [20]–[22], [47]. All these experiments demonstrated that quantum mechanical laws hold true: experimental probabilities coincide (of course, with some precision) with quantum probabilities $\mathbf{P}_q(\mathbf{s}_\theta = \epsilon, \mathbf{s}_\phi = \delta), \epsilon, \delta = \pm 1$, computed via (6.3), (6.4). Bell's inequality for experimental probabilities is violated.

1. Probability and reality. It is widely accepted by a part of physical community that the violation of Bell's inequality has demonstrated that the realists philosophy cannot be used for the description of quantum phenomena: the spin is not an objective property of a quantum particle.

Remark 7.1. Other part of the physical community connects Bell's inequality and nonlocality of space-time: in the real physical experiments observables $\mathbf{s}_\theta = \mathbf{s}_\theta^1$ and $\mathbf{s}_\phi = \mathbf{s}_\phi^2$ correspond to two particles which are separated in space. However, our probabilistic analysis will demonstrate that there are no traces of nonlocality in Bell's framework. Therefore we will mainly concentrate our considerations on connection between Bell's inequality and realism.

The problem of existence (reality) of spin is often mixed with the problem of existence of random variables $\mathbf{s}_\phi(\omega), [0, 2\pi]$, defined on some Kolmogorov probability space. However, these are two different problems. Kolmogorov's model is just one of possible models of reality. Besides Kolmogorov's model, there exist frequency and ensemble models. We shall demonstrate that Bell's inequality does not present in the latter models. Thus there is no problem with experimental violations of this inequality. The spin can be in principle considered as an objective property of a quantum system. We shall show that we can even keep to 'real realism', namely, i-realism.

2. Realism and Bell's inequality. Let $s_\phi, \phi \in [0, 2\pi]$, be initial values. We use the ensemble approach to probability. The main distinguishing feature of this approach is that all probabilities in (6.5)–(6.7) depend on corresponding ensembles. There are three ensembles $S_{0\phi}^N, S_{\phi\theta}^N, S_{0\theta}^N$ (of cardinality N) which are used to obtain observed probabilities $\mathbf{P}_{S_{0\phi}^N}(s_0 = \pm 1, s_\phi = \pm 1), \mathbf{P}_{S_{\phi\theta}^N}(s_\phi = \pm 1, s_\theta = \pm 1), \mathbf{P}_{S_{0\theta}^N}(s_0 = \pm 1, s_\theta = \pm 1)$.[19]

[19]In Kolmogorov's model ensemble indexes are omitted. In fact, this manipulation which

Remark 7.2. Formally we could introduce an infinite ensemble S of particles and define ensemble probabilities with respect to S :

$$\mathbf{P}_S(s_{\alpha_1}=\epsilon_1, s_{\alpha_2}=\epsilon_2, ..., s_{\alpha_n}=\epsilon_n)=\frac{|\{s \in S : s_{\alpha_1}=\epsilon_1, s_{\alpha_2}=\epsilon_2, ..., s_{\alpha_n}=\epsilon_n\}|}{|S|}, \tag{7.1}$$

where $\epsilon_i=\pm 1, n \in \mathbf{N}$. Of course, the calculations of section 6 can be repeated for such probabilities. However, this can be done only formally. As we have already mentioned, the proportional definition of probability is meaningless for infinite ensembles in the framework of real analysis. Thus we could not perform these formal calculations on the mathematical level of rigorousness.

Remark 7.3. The proportional ensemble definition (7.1) can be used on the mathematical level of rigorousness on the basis of non-Archimedean analysis. For example, in Chapter 4 we study p-adic ensemble probabilities. All arithmetical calculations (6.5)- (6.9) can be performed in the field of p-adic numbers. But (6.9) does not imply (6.10)! Some of probabilities $\mathbf{P}_S(s_{\alpha_1}=\epsilon_1, s_{\alpha_2}=\epsilon_2, s_{\alpha_3}=\epsilon_3)$ can be negative! In fact, there is some hidden (and still unclear) logic in such an appearance of negative probabilities in models in that the formal use of infinite statistical ensembles is not justified (see Chapter 3).

Therefore we have to operate with finite ensembles $S^N_{\alpha\beta}$. The three dimensional probabilities used in (6.5)–(6.7) must be also considered as probabilities with respect to these ensembles. Thus in (6.5)– (6.7) we use probabilities $\mathbf{P}_{S^N_{0\phi}}(...), ..., \mathbf{P}_{S^N_{0\theta}}(...)$. Hence (6.5)+(6.6) gives

$$\mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0=+1, \mathbf{s}_\phi=+1)+\mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_\phi=-1, \mathbf{s}_\theta=+1)$$

$$=\mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0=+1, \mathbf{s}_\phi=+1, \mathbf{s}_\theta=+1)+\mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0=+1, \mathbf{s}_\phi=+1, \mathbf{s}_\theta=-1)$$

$$+\mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_0=+1, \mathbf{s}_\phi=-1, \mathbf{s}_\theta=+1)+\mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_0=-1, \mathbf{s}_\phi=-1, \mathbf{s}_\theta=+1).$$

But in the opposite to calculations with abstract Kolmogorov probabilities in section 6 the first and the third terms on the right hand side of this equation are not those which when added together make up the term

$$\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0=+1, \mathbf{s}_\theta=+1)=$$

$$=\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0=+1, \mathbf{s}_\phi=+1, \mathbf{s}_\theta=+1)+\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0=+1, \mathbf{s}_\phi=-1, \mathbf{s}_\theta=+1).$$

looks quite innocent is the origin of Bell's mystification.

To obtain (6.9), we have to identify $\mathbf{P}_{S^N_{0\phi}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1, \mathbf{s}_\theta = +1)$ and $\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = +1, \mathbf{s}_\theta = +1)$, $\mathbf{P}_{S^N_{\phi\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1)$ and $\mathbf{P}_{S^N_{0\theta}}(\mathbf{s}_0 = +1, \mathbf{s}_\phi = -1, \mathbf{s}_\theta = +1)$. But (and this is the crucial point!) there are no reasons to do this in the general case.

For example, in quantum experiments with correlated particles it is possible to measure only two dimensional probabilities $\mathbf{P}_{S^N_{\alpha_1\alpha_2}}(s_{\alpha_1} = \pm 1, s_{\alpha_2} = \pm 1)$ (by using correlations between particles). The physical experience is that this ensemble probabilities stabilize when $N \to \infty$. However, there are no reasons that three dimensional probabilities $\mathbf{P}_{S^N_{\alpha_1\alpha_2}}(s_{\alpha_1} = \pm 1, s_{\alpha_2} = \pm 1, s_{\alpha_3} = \pm 1)$ must also stabilize when $N \to \infty$. They could depend essentially on statistical ensembles. Therefore the identification of probabilities with respect to different ensembles is not justified at all.

Practically the same considerations can be repeated in the framework of von Mises' frequency theory. There we have to consider three different collectives, $x_{0\phi}, x_{\phi\theta}, x_{0\theta}$, instead of ensembles $S^N_{0\phi}, S^N_{\phi\theta}, S^N_{0\theta}$. These are collectives for two dimensional labels $(s_0, s_\phi), (s_\phi, s_\theta), (s_0, s_\theta)$. The principle of statistical stabilization can be applied only to these lables. The frequencies $\nu_N(s_\alpha = \pm 1, s_\beta = \pm 1; x_{\alpha\beta})$ stabilize when $N \to \infty$. However, the frequencies $\nu_N(s_\alpha = \pm 1, s_\beta = \pm 1, s_\gamma = \pm 1; x_{\alpha\beta})$ need not stabilize when $N \to \infty$. Moreover, they may be not defined at all. Therefore *there is no Bell's inequality in von Mises probability model.*

If we keep to f-realism, we have to use von Mises' frequency probability theory. Therefore we could not obtain Bell's inequality. There are no problems with violations of this inequality.

Remark 7.4. Typically Bell's inequality is associated with the use of so called hidden variables, see section 9. As it has been noticed in [30], [33], it can be derived without any reference to hidden variables. As the reader has seen, it was only supposed that there exists a Kolmogorov probability space $\mathcal{P} = \{\Omega, \mathcal{F}, \mathbf{P})$ such that three spin projections $\mathbf{s}_0, \mathbf{s}_\phi, \mathbf{s}_\theta$ can be represented by random variables on this space. Under this assumption it is possible to define the joint probability distribution $\mathbf{P}_{ijk} = \mathbf{P}(\mathbf{s}_0 = i, \mathbf{s}_\phi = j, \mathbf{s}_\theta = k), i, j, k = \pm 1$. On the other hand, the existence of the joint probability distribution $\mathbf{P}_{ijk}$ implies the existence of the Kolmogorov space with $\mathbf{P} = \{\mathbf{P}_{ijk}\}$. This connection between Bell's inequality and existence of joint probability distribution was discussed by A. Fine [38], P. Rastall [95], W. de Muynck and H. Martens [32] (see also [33]). Typically nonexistence of the joint probability distribution is interpreted as the impossibility to use the objective realism (at least its i-version). From our viewpoint this is just the impossibility to

apply the Kolmogorov model of probability theory (i.e., to use abstract symbolic probabilities without to regard to concrete ensembles or collectives). In the frequency approach such nonexistence only demonstrates the absence of the statistical stabilization for relative frequencies $\nu_N(\mathbf{s}_0 = i, \mathbf{s}_\phi = j, \mathbf{s}_\theta = k)$ for three different projections of spin. However, there are no (!) experimental reasons to suppose such a stabilization. In the ensemble approach such nonexistence only demonstrates the absence of the reproducibility of the 'property' $(\mathbf{s}_0 = i, \mathbf{s}_\phi = j, \mathbf{s}_\theta = k)$ in statistical ensembles used for quantum experiments. However, there are no(!) reasons to suppose such a reproducibility.

8 Bell's inequality for covariations

We have considered Bell's inequality for probabilities. The original Bell's inequality [10], [11] was proved for covariations.

Theorem 8.1. *Let $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ be a Kolmogorov probability space and $A, B, C \in RV(\mathcal{P})$ be discrete random variables, $A, B, C = \pm 1$. Then Bell's inequality*

$$| < A, B > - < C, B > | \leq 1 - < A, C > \tag{8.1}$$

holds true.

Proof. Set $\Delta = < A, B > - < C, B >$. By linearity of Lebesgue integral we obtain

$$\Delta = \int_\Omega A(\omega)B(\omega)d\mathbf{P}(\omega) - \int_\Omega C(\omega)B(\omega)d\mathbf{P}(\omega) = \int_\Omega [A(\omega) - C(\omega)]B(\omega)d\mathbf{P}(\omega). \tag{8.2}$$

As $A(\omega)^2 = 1$,

$$|\Delta| = | \int_\Omega [1 - A(\omega)C(\omega)]A(\omega)B(\omega)d\mathbf{P}(\omega)| \tag{8.3}$$

$$\leq \int_\Omega [1 - A(\omega)C(\omega)]d\mathbf{P}(\omega) = 1 - < A, C > . \tag{8.4}$$

■

Of course, this is the rigorous mathematical proof of (8.1) for Kolmogorov probabilities. However, as we have mentioned, Kolmogorov's model does not provide the adequate description of some quantum measurements. The root of 'Bell-Kolmogorov mystification' is again the identification of probabilities corresponding to different statistical ensembles or collectives.

Let us consider von Mises' approach. The ensemble approach will be considered in section 9 in connection with so called hidden variables. In the frequency formalism the covariations $< A, B >$ and $< C, B >$ are covariations with respect to two different collectives, x_{AB} and x_{CB} : $< A, B > \equiv < A, B >_{x_{AB}}$ and $< C, B > \equiv < C, B >_{x_{CB}}$. Thus

$$< A, B >_{x_{AB}} = \frac{1}{N} \sum_{i=1}^{N} a_i b_i \quad , \quad < C, B >_{x_{CB}} = \frac{1}{N} \sum_{i=1}^{N} c_i b'_i$$

and

$$< A, B > x_{AB} - < C, B > x_{CB} = \frac{1}{N} \sum_{i=1}^{N} [a_i b_i - c_i b'_i].$$

There are no reasons to suppose that

$$\frac{1}{N} \sum_{i=1}^{N} [a_i b_i - c_i b'_i] = \frac{1}{N} \sum_{i=1}^{N} [a_i - c_i] b_i. \tag{8.5}$$

Hence Bell made the mistake on the first step, (8.2), of the proof by using the linearity of mean value with respect to two different collectives (or statistical ensembles).

As physicists (with a few exceptions) did not see probabilistic roots of Bell's misunderstanding, they try to find some explanations of experimental violations of Bell's inequality:

1. **Death of reality.** It is impossible to keep to realism and suppose that quantum systems have objective properties.

2. **Nonlocality.** As in quantum experiments, covariations are found via measurements for correlated particles which are separated in space, it can be supposed that 'nonclassical' behaviour of these covariations is a consequence of the dependence of the state of one particle on the state of other particle.

Of course, these ideas could not be denied on the basis of our probability analysis. But our analysis has demonstrated that Bell's arguments have no relation to these ideas.

9 Hidden variables and Bell's inequality

1. Incompleteness of quantum mechanics. Theories based on so called *hidden variables* were developed starting with the hypothesis on incompleteness of quantum mechanics. Typically incompleteness of quantum mechanics

is considered as a consequence of the EPR arguments. By these arguments both position and momentum (or projections of spin to different axes) of each quantum particle are elements of reality for the system of two correlated particles. However, the quantum formalism does not describe the simultaneous existence of these elements of reality. Thus quantum mechanics is not complete. However, we have demonstrated that the EPR arguments are based on the formal use of Kolmogorov abstract probabilities. Of course, such 'arguments' could not be considered as the proof of incompleteness of quantum mechanics. Nevertheless, incompleteness of quantum mechanics can be directly obtained as a consequence of keeping to realist philosophy.

2. Hidden variables. Let us suppose that quantum mechanics is not complete. There could be finer description of reality than given by quantum mechanics. In principle there could exist some additional variables λ, hidden variables, such that by specifying the value $\lambda = \lambda_0$ of λ we could determine the values of all physical observables: $\lambda_0 \to A(\lambda_0), \lambda_0 \to B(\lambda_0), \ldots.$. Compatibility or incompatibility of physical observables $A, B, \ldots$, do not play any role.

Typically Bell's inequality is considered in the framework of hidden variables. The Kolmogorov probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$ which was used in section 8 has the following interpretation: $\Omega = \Lambda$ is the set of hidden variables, $\omega = \lambda$, $\mathbf{P}$ is the probability distribution of hidden variables. The experimental violations of Bell's inequality are interpreted as the evidence that such *hidden variables do not exists.* Other authors use nonlocality arguments. They think that, despite of nonexistence of local hidden variables, *nonlocal hidden variables* can exist.

However, all our probability arguments against Bell's inequality can be repeated for hidden variables. Let us use von Mises' frequency approach. As we have already seen in section 8, Bell's mistake is the assumption on the validity of equality (8.5).

3. Deterministic hidden variables model and Bell's mystification. To simplify our considerations, we suppose that the set of hidden variables is finite:

$$\Lambda = \{\lambda_1, \ldots, \lambda_M\}.$$

For each physical observable U, the value λ of hidden variables determines the value

$$U = U(\lambda).$$

Here we keep to realism. It is possible to keep i-realism or f-realism. If

we keep to i-realism in this model, we have to assume that the result of measurement does not depend on fluctuations of an internal state ω of a measurement device $\mathcal{M}_U$ (see the next section for the model with such a dependence).

Let U and V be physical observables, $U, V = \pm 1$. We start with the consideration of the frequency (experimental) covariation $< U, V >_{x_{UV}}$ with respect to a collective x_{UV} induced by measurements of the pair (U, V). The x_{UV} is obtained by measurements for an ensemble S_{UV} of physical systems (for example, pairs of correlated quantum particles). Our aim is to represent experimental covariation $< U, V >_{x_{UV}}$ as ensemble covariation $< U, V >_{S_{UV}}$. Then we shall demonstrate that in the general case it is impossible to perform for ensemble covariations Bell's calculations, (8.2) – (8.4), which have been performed for Kolmogorov covariations. Let $S_{UV} = \{d_1, ..., d_N\}$, where ith measurement is performed for the system d_i. Define a function $i \to \lambda(i)$, the value of hidden variables for d_i. We set $n_k(S_{UV}) = |\{d_i \in S_{UV} : \lambda(i) = \lambda_k\}|$ and $\mathbf{p}_k^{UV} = \mathbf{P}_{S_{UV}}(\lambda = \lambda_k) = \frac{n_k(S_{UV})}{N}$. We have

$$< U, V >_{x_{UV}} = \frac{1}{N} \sum_{i=1}^{N} U(\lambda(i)) V(\lambda(i)) = \frac{1}{N} \sum_{k=1}^{M} n_k(S_{UV}) u_k v_k$$

$$= \sum_{k=1}^{M} \mathbf{p}_k^{UV} u_k v_k = < U, V >_{S_{UV}},$$

where $u_k = U(\lambda_k), v_k = V(\lambda_k)$. Thus

$$\Delta = < A, B >_{x_{AB}} - < C, B >_{x_{CB}}$$

$$= < A, B >_{S_{AB}} - < C, B >_{S_{CB}} = \sum_{k=1}^{M} (\mathbf{p}_k^{AB} a_k - \mathbf{p}_k^{CB} c_k) b_k$$

and

$$< A, C >_{x_{AC}} = < A, C >_{S_{AC}} = \sum_{k=1}^{M} \mathbf{p}_k^{AC} a_k c_k.$$

We suppose now that *probabilities of* λ_k *do not depend on statistical ensembles:*

$$\mathbf{p}_k = \mathbf{p}_k^{AB} = \mathbf{p}_k^{CB} = \mathbf{p}_k^{AC} \tag{9.1}$$

(later we shall modify this condition to obtain statistical coincidence of probabilities, instead of the precise coincidence). Hence

$$\Delta = \sum_{k=1}^{M} \mathbf{p}_k (a_k - c_k) b_k \quad \text{and} \quad < A, C >_{x_{AC}} = \sum_{k=1}^{M} \mathbf{p}_k a_k c_k.$$

We can now apply Theorem 8.1 for the discrete probability distribution $\{\mathbf{p}_k\}_{k=1}^{M}$ and obtain Bell's inequality.

However, if condition (9.1) does not hold true, then equality (8.2) and, as a consequence, Bell's inequality can be violated. The violation of condition (9.1) is the exhibition of unstable (with respect to the real metric) statistical structure on the level of hidden variables of (at least some) quantum ensembles. In particular, the principle of the statistical stabilization ('the law of the large numbers') can be violated for hidden variables: $\lim_{N\to\infty} \nu_N(\lambda_l)$ do not exist. Thus we could not introduce the probability distribution on the set of hidden labels Λ. [20]

Nevertheless, we obtained the following mathematical result:

Theorem 9.1. *Let statistical ensembles satisfy condition (9.1). Then Bell's inequality (8.1) holds true.*

We introduce now a statistical analogue of the precise coincidence of ensemble probabilities for hidden variables. Let $\mathcal{E}_1, \mathcal{E}_2$ be two ensembles of physical systems and let π be a property of elements of these ensembles. The π has values $(\alpha_1, ..., \alpha_m)$. We define

$$\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) = \sum_{i=1}^{M} |\mathbf{P}_{\mathcal{E}_1}(\alpha_i) - \mathbf{P}_{\mathcal{E}_2}(\alpha_i)|,$$

where $\mathbf{P}_{\mathcal{E}}(\alpha_i) = \frac{|\{d\in\mathcal{E}:\pi(d)=\alpha_i\}|}{|\mathcal{E}|}$ are ensemble probabilities. We remark that the function δ_π is a pseudometric on the set of all ensembles which elements have the property π : (1) $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) \geq 0$; (2) $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) = \delta_\pi(\mathcal{E}_2, \mathcal{E}_1)$; (3) $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) \leq \delta_\pi(\mathcal{E}_1, \mathcal{E}_3) + \delta_\pi(\mathcal{E}_3, \mathcal{E}_2)$. The distance $\delta_\pi(\mathcal{E}_1, \mathcal{E}_2) = 0$ iff ensembles $\mathcal{E}_1$ and $\mathcal{E}_2$ have the same probability distribution of property $\pi : \mathbf{P}_{\mathcal{E}_1}(\alpha_i) = \mathbf{P}_{\mathcal{E}_2}(\alpha_i), i = 1, 2, ..., m.$

---

[20] All our considerations were based on the statistical stabilization with respect to the real metric. In Chapter 4 we shall consider the statistical stabilization with respect to a p-adic metric. It may be that some ensembles of hidden variables which are unstable with respect to the real metric are stable with respect to the p-adic metric, see [58]–[60].

In our model we set $\pi = \lambda$, hidden variables. The precise repeatability of the probability distribution of hidden variables (9.1) can be written as

$$\delta(S_{AB}, S_{CB}) = \delta(S_{AB}, S_{AC}) = 0,$$

where $\delta = \delta_\lambda$. Of course, we need not use such a precise coincidence in probabilistic considerations.

Theorem 9.2. *Let statistical ensembles satisfy condition*

$$\delta(S_{AB}, S_{CB}), \delta(S_{AB}, S_{AC}) \le \epsilon.$$

Then Bell's inequality

$$| < A, B >_{S_{AB}} - < C, B >_{S_{CB}} | \le (1 + 2\epsilon) - < A, C >_{S_{AC}} \tag{9.2}$$

holds true.

Proof. We have

$$|\Delta| \le |\sum_{k=1}^{M} \mathbf{p}_k^{AB}(a_k - c_k)b_k| + |\sum_{k=1}^{M}(\mathbf{p}_k^{AB} - \mathbf{p}_k^{CB})c_k b_k|$$

$$\le \epsilon + \sum_{k=1}^{M} \mathbf{p}_k^{AB}|a_k b_k|(1 - a_k c_k) \le (1+\epsilon) - < A, C >_{S_{AC}} + \sum_{k=1}^{M} |\mathbf{p}_k^{AC} - \mathbf{p}_k^{AB}||a_k c_k|$$

$$\le (1 + 2\epsilon) - < A, C >_{S_{AC}} .$$

■

We use the index N to denote the cardinality of a statistical ensemble. If probabilities $\mathbf{P}_{S_{UV}^N}(\lambda_k)$ stabilize when $N \to \infty$,

$$\lim_{N\to\infty} \mathbf{P}_{S_{UV}^N}(\lambda_k) = \mathbf{P}(\lambda_k),$$

then $\delta(S_{AB}, S_{CB}), \delta(S_{AB}, S_{AC}) \to 0, N \to \infty$. This imply precise Bell's inequality (8.1). On one hand, experimental violations of the latter inequality can demonstrate that *probabilities of hidden variables with respect to the ideal infinite ensemble do not exist at all (they fluctuate when $N \to \infty$).* On the other hand, these violations can be a consequence of the fact that we do not know the value of a constant ϵ in (9.2). It might be that, despite of the stabilization of probabilities for $N \to \infty$, this constant is quite large for

statistical ensembles which are used in quantum physics. In fact, 'right Bell's inequality' (9.2) could not be experimentally verified.

4. Stochastic hidden variables model and Bell's mystification. Here we keep to f-realism. Thus, for each physical observable U, its value $U_f = U(\lambda)$ is the final value of U after a measurement. Such a result of a measurement depends not only on the value λ of hidden variables, but also on the state of an equipment $\mathcal{M}_U$ which is used for measuring of U. A measurement device $\mathcal{M}_U$ is a complex macroscopic system which state depends on the huge number of fluctuating parameters. Denote the ensemble of all possible states of $\mathcal{M}_U$ by the symbol Σ_U : $\Sigma_U = \{\omega_1^U, ..., \omega_{L_U}^U\}$. The final value U_f of an observable U depends on both λ and ω :

$$U_f = U(\omega, \lambda).$$

We call such a model *stochastic hidden variables model.* Our definition of a stochastic hidden variables model differs from the standard one, see subsection 5. The standard definition is strongly connected with Kolmogorov's model.

Let U and V be physical observables, $U, V = \pm 1$. We start again with the consideration of the frequency covariation $< U, V >_{x_{UV}}$ with respect to a collective x_{UV} induced by the measurement of the pair (U, V). The x_{UV} is obtained by measurements for an ensemble S_{UV} of physical systems. Our aim is again to represent the experimental covariation $< U, V >_{x_{UV}}$ as ensemble covariation $< U, V >_{S_{UV}}$. Then we shall demonstrate that in the general case it is impossible to perform for ensemble covariations Bell's calculations, (8.2) - (8.4).

Let $S_{UV} = \{d_1, ..., d_N\}$, where ith measurement is performed for the system d_i. Define functions $i \to \lambda(i)$ (the same function as above) and $i \to \omega^U(i), i \to \omega^V(i)$, states of apparatus $\mathcal{M}_U$ and $\mathcal{M}_V$, respectively, at the instances, t_i^U and t_i^V, of measurements of U and V for ith system. We have

$$< U, V >_{x_{UV}} = \frac{1}{N} \sum_{i=1}^{N} U(\omega^U(i), \lambda(i)) V(\omega^V(i), \lambda(i)).$$

Set $D_{ks}^U = \{i : \lambda(i) = \lambda_k, \omega^U(i) = \omega_s^U\}$ and $D_{ks}^V = \{i : \lambda(i) = \lambda_k, \omega^V(i) = \omega_s^V\}, 1 \leq k \leq M, 1 \leq s \leq L_U, 1 \leq q \leq L_V$. Set $l_{ksq}^{UV} = |D_{ks}^U \cap D_{kq}^V|$. It is evident that

$$\sum_{k=1}^{M} \sum_{s=1}^{L_U} \sum_{q=1}^{L_V} l_{ksq}^{UV} = N.$$

Hence

$$< U, V >_{x_{UV}} = \frac{1}{N} \sum_{ksq} l_{ksq}^{UV} u_{ks} v_{kq},$$

where $u_{ks} = U(\omega_s^U, \lambda_k), v_{kq} = V(\omega_q^V, \lambda_k)$. We show that $< U, V >_{x_{UV}}$ can be represented as ensemble covariation for an appropriative ensemble of physical systems and states of measurement devices.

First we note that $< U, V >_{x_{UV}} \neq < U, V >_{\Lambda \times \Sigma_A \times \Sigma_B}$ (compare with subsection 5). For the latter covariation, we have

$$< U, V >_{\Lambda \times \Sigma_A \times \Sigma_B} = \frac{1}{M L_A L_B} \sum_{k=1}^{M} \sum_{s=1}^{L_U} \sum_{q=1}^{L_V} u_{ks} v_{kq}$$

and in general $\mathbf{P}_{\Lambda \times \Sigma_A \times \Sigma_B}(\lambda = \lambda_k, \omega^U = \omega_s^U, \omega^V = \omega_q^V) = \frac{1}{M L_A L_B} \neq \frac{l_{ksq}}{N}$ even approximately for $M, N, L_A, L_B \to \infty$.

It is also evident that $< U, V >_{x_{UV}} \neq < U, V >_{S_{UV}}$. The latter covariation is simply not well defined, because the 'properties' $\omega^U(i) = \omega_s^U, \omega^V(i) = \omega_q^V$ are not objective properties of elements of the ensemble S_{UV}. These 'properties' are determined by fluctuations of parameters in the apparatus $\mathcal{M}_U$ and $\mathcal{M}_V$.

To find the right ensemble, we have to introduce two new ensembles, namely, ensembles of states of the apparatus $\mathcal{M}_U$ and $\mathcal{M}_V$ (in the process of measurements for the ensemble of physical systems S_{UV}) :

$$S_{\mathcal{M}_U} = \{\alpha_1^U, ..., \alpha_N^U\}, \; \alpha_j^U \in \Sigma_U, \;\; S_{\mathcal{M}_V} = \{\alpha_1^V, ..., \alpha_N^V\},, \alpha_j^V \in \Sigma_V,$$

where $\alpha_i^U = \omega^U(i), \alpha_i^V = \omega^V(i)$ are states of $\mathcal{M}_U$ and $\mathcal{M}_V$ at the instances of ith measurements. We set

$$\mathbf{S}_{UV} = \mathrm{diag}(\mathrm{S_{UV}} \times \mathrm{S}_{\mathcal{M}_\mathrm{U}} \times \mathrm{S}_{\mathcal{M}_\mathrm{V}}) = \{\mathrm{D}_1, ..., \mathrm{D_N}\}, \;\; \mathrm{D_j} = (\mathrm{d_j}, \alpha_\mathrm{j}^\mathrm{U}, \alpha_\mathrm{j}^\mathrm{V}).$$

Then $\pi(D_j) = (\lambda(j), \omega^U(j), \omega^V(j))$ is an objective property of elements of the ensemble $\mathbf{S}_{UV}$ and

$$< U, V >_{x_{UV}} = < U, V >_{\mathbf{S}_{UV}} = \frac{1}{N} \sum_{i=1}^{N} U(\omega^U(i), \lambda(i)) V(\omega^V(i), \lambda(i)).$$

We set

$$\mathbf{p}_{ksq}^{UV} = \mathbf{P}_{\mathbf{S}_{UV}}(D_j : \pi(D_j) = (\lambda_k, \omega_s^U, \omega_s^V))$$

$$= \frac{|\{D_j \in \mathbf{S}_{UV} : \pi(D_j) = (\lambda_k, \omega_s^U, \omega_s^V)\}|}{|\mathbf{S}_{UV}|}.$$

Hence we obtained that

$$< U, V >_{x_{UV}} = < U, V >_{\mathbf{S}_{UV}} = \sum_{ksq} \mathbf{p}_{ksq}^{UV}\ u_{ks} v_{kq}.$$

Thus in the general case we have

$$\Delta = < A, B >_{x_{AB}} - < C, B >_{x_{CB}} = < A, B >_{\mathbf{S}_{AB}} - < C, B >_{\mathbf{S}_{CB}}$$

$$= \sum_{ksq} \mathbf{p}_{ksq}^{AB}\ a_{sk} b_{kq} - \sum_{ksq} \mathbf{p}_{ksq}^{CB}\ c_{ks} b_{kq}$$

and

$$< A, C >_{x_{AC}} = < A, C >_{\mathbf{S}_{AC}} = \sum_{ksq} \mathbf{p}_{ksq}^{AC}\ a_{ks} c_{kq}.$$

We suppose now that probabilities $\mathbf{p}_{ksq}^{UV}$ do not depend on ensembles:

$$\mathbf{p}_{ksq} = \mathbf{p}_{ksq}^{AB} = \mathbf{p}_{ksq}^{CB} = \mathbf{p}_{ksq}^{AC}. \tag{9.3}$$

In particular, we suppose that all measurement devices have the same set of states (of parameters):

$$\Sigma = \Sigma_A = \Sigma_B = \Sigma_C \ \ (\text{and } \ L = L_A = L_B = L_C). \tag{9.4}$$

Then we obtain

$$\Delta = \sum_{ksq} \mathbf{p}_{ksq}(a_{ks} - c_{ks}) b_{kq}.$$

However, we could not repeat trick (8.3) of the proof of Bell's inequality. The equality $a_{ks}^2 = 1$ does not give the possibility to proceed the proof. Of course, we have

$$|\Delta| = |\sum_{ksq} \mathbf{p}_{ksq}(a_{ks} - a_{ks}^2 c_{ks}) b_{kq}| \le \sum_{ksq} \mathbf{p}_{ksq} |a_{ks} b_{kq}| (1 - a_{ks} c_{ks})$$

$$\le 1 - \sum_{ksq} \mathbf{p}_{ksq} a_{ks} c_{ks}.$$

But in general $\sum_{ksq} \mathbf{p}_{ksq} a_{ks} c_{ks}$ is not larger than $< A, C >_{x_{AC}} = \sum_{ksq} \mathbf{p}_{ksq} a_{ks} c_{kq}$.

Therefore, if we keep to f-realism, even stability condition (9.3) (for combined ensembles of physical systems and states of measurement apparatus)

does not imply Bell's inequality. A new source of violation of Bell's inequality is the *inconsistency* of random fluctuations for two measurement devices $\mathcal{M}_U$ and $\mathcal{M}_V$. In general $\omega^U(i) \neq \omega^V(i)$.

Suppose that it could be possible to control states of $\mathcal{M}_U$ and $\mathcal{M}_V$ and choose ω for $\mathcal{M}_U$ and $\mathcal{M}_V$ in the consistence way:

$$\omega = \omega^U(i) = \omega^V(i).$$

Then the ensemble $\mathbf{S}_{UV}$ would contain only triples of the form $(\lambda_k, \omega_s, \omega_s)$ and

$$\mathbf{p}_{ksq}^{UV} = \mathbf{P}_{\mathbf{S}_{UV}}(\lambda_k, \omega_s^U, \omega_q^V) = 0, \; s \neq q. \tag{9.5}$$

In such a case we obtain covariations:

$$< U, V >_{\text{Ideal}} = \frac{1}{N} \sum_{i=1}^{N} U(\omega^U(i), \lambda(i)) V(\omega^V(i), \lambda(i)) = \sum_{ks} \mathbf{p}_{ks}^{UV} u_{ks} v_{ks},$$

where $\mathbf{p}_{ks}^{UV} = \mathbf{p}_{kss}^{UV}$. If we also suppose the validity of (9.3), we obtain

$$|\Delta_{\text{Ideal}}| = |\sum_{ks} \mathbf{p}_{ks}(a_{ks} - c_{ks}) b_{ks}|$$

$$\leq 1 - \sum_{ks} \mathbf{p}_{ks} a_{ks} c_{ks} = 1 - < A, C >_{\text{Ideal}} .$$

However, ideal covariations have no direct connection to experimental frequency covariations.

Nevertheless, we can formulate the following mathematical theorem:

Theorem 9.3. *Let statistical ensembles (physical systems/measurement apparatus) satisfy conditions (9.3) and (9.5). Then Bell's inequality (8.1) holds true for covariations with respect to these ensembles.*

Therefore, to obtain Bell's inequality in the framework of f-realism, we have to suppose: (1) statistical repeatability of ensemble distribution of hidden variables λ in ensembles which are used for measurements; (3) statistical repeatability of fluctuations of states ω in ensembles of an equipment; (3) consistency of fluctuations of all measurement devices.

If the reader even deny the possibility of violations of (1) or (2), he must agree that condition (3) seems to be nonphysical: we could never control fluctuations of the huge number of parameters in the equipment.

Instead of precise coincidence (9.3), it is possible to consider (under the assumption (9.4)) the statistical coincidence based on the quantity:

$$\delta(\mathbf{S}_{AB}, \mathbf{S}_{CB}) = \sum_{k=1}^{M} \sum_{s=1}^{L} \sum_{q=1}^{L} |\mathbf{p}_{ksq}^{AB} - \mathbf{p}_{ksq}^{CB}|.$$

Here $\delta = \delta_\pi$ for the property $\pi(i) = (\lambda(i), \omega^U(i), \omega^V(i))$. We remark that condition (9.3) of the precise coincidence can be written as

$$\delta(\mathbf{S}_{AB}, \mathbf{S}_{CB}) = 0$$

for every two pairs of observable (A, B) and (C, B). We also introduce a new quantity which is a statistical measure of inconsistency of ensembles $S_{\mathcal{M}_U}$ and $S_{\mathcal{M}_V}$:

$$\sigma(\mathbf{S}_{UV}) = \sum_{s \neq q} \mathbf{P}_{\mathbf{S}_{UV}}(\omega^U = \omega_s, \omega^V = \omega_q) = \sum_k \sum_{s \neq q} \mathbf{p}_{ksq}^{UV}.$$

Condition (9.5) of the precise consistency for states of $\mathcal{M}_U$ and $\mathcal{M}_V$ can be written in the form:

$$\sigma(\mathbf{S}_{UV}) = 0.$$

Theorem 9.4. *Let statistical ensembles (physical systems/measurement apparatus) satisfy conditions:*

$$\delta(\mathbf{S}_{AB}, \mathbf{S}_{CB}), \delta(\mathbf{S}_{AB}, \mathbf{S}_{AC}) \leq \epsilon \ \textit{ and } \ \sigma(\mathbf{S}_{AB}), \sigma(\mathbf{S}_{CB}), \sigma(\mathbf{S}_{AC}) \leq \epsilon'.$$

Then inequality

$$| < A, B >_{\mathbf{S}_{AB}} - < C, B >_{\mathbf{S}_{CB}} | \leq (1 + 2\epsilon + 3\epsilon') - < A, C >_{\mathbf{S}_{AC}}$$

holds true.

Proof. We have

$$|\Delta| \leq \epsilon + |\sum_{ksq} \mathbf{p}_{ksq}^{AB}(a_{ks} - c_{ks})b_{kq}|$$

$$\leq \epsilon + 2\epsilon' + \sum_{ks} \mathbf{p}_{ks}^{AB}|(a_{ks} - c_{ks})b_{ks}| \leq \epsilon + 2\epsilon' + \sum_{ks} \mathbf{p}_{ks}^{AB}(1 - a_{ks}c_{ks})$$

$$\leq \epsilon + 4\epsilon' + \sum_{ksq} \mathbf{p}_{ksq}^{AB}(1 - a_{ks}c_{kq}) \leq (1 + 2\epsilon + 4\epsilon') - \sum_{ksq} \mathbf{p}_{ksq}^{AC} a_{ks}c_{kq}.$$

■

5. Right choice of probability distributions for stochastic hidden variables models. Typically stochastic hidden variables models are defined as models with probabilities ($\epsilon = \pm 1$)

$$\mathbf{P}(U = \epsilon) = \int_\Lambda \mathbf{P}(U = \epsilon/\lambda) d\, \rho(\lambda), \tag{9.6}$$

where $\rho(\lambda)$ is the probability distribution of hidden variables and $\mathbf{P}(U = \epsilon/\lambda)$ is the conditional probability to measure the value $U = \epsilon$ for the quantum system having the hidden state λ. Then the joint probability distribution can be defined (at least mathematically) as

$$\mathbf{P}(U_1 = \epsilon_1, U_2 = \epsilon_2, U_3 = \epsilon_3) = \int_\Lambda \mathbf{P}(U_1 = \epsilon_1/\lambda)\mathbf{P}(U_2 = \epsilon_2/\lambda)\mathbf{P}(U_3 = \epsilon_3/\lambda) d\, \rho(\lambda). \tag{9.7}$$

In fact, to derive Bell's inequality in the Kolmogorov framework, it is sufficient to use the existence (on the mathematical level) of the joint probability distribution (9.7). However, considerations in the framework of the ensemble probability theory demonstrated that 'probabilities' (9.6) has no physical meaning. These are probabilities with respect to the ensemble $\Lambda \times \Sigma_U$. However, physical probabilities are probabilities with respect to the ensemble $\mathbf{S}_U = \mathrm{diag}(\mathrm{S_U} \times \mathrm{S}_{\mathcal{M}_\mathrm{U}})$, where $S_U = \{d_1, ..., d_N\}$ is the ensemble of quantum system used in the measurement.

6. Individual and ensemble nonreproducibilities. Our hypothesis on the nonreproducibility of the probability distribution of hidden variables in statistical ensembles used in quantum experiments is related to De Baere's [25], [26] hypothesis on the *individual nonreproducibility.* He mentioned that there are reasons (see also [33]) that it would be impossible to prepare the quantum system with the same value λ for measurements of different observables. Thus probabilities $\mathbf{P}(U_j = \epsilon_j/\lambda), j = 1, 2, 3,$ could not be defined for the same λ. The latter implies that joint probability distribution (9.7) does not exist and Bell's inequality could not be derived. We remark that deterministic hidden variables models satisfy the condition of the individual reproducibility. However, the ensemble reproducibility can be violated.

7. Contextualistic realism. All our considerations in the f-realist framework can be repeated in the framework of so called contextualistic realism (N. Bohr and W. Heisenberg, see, for example, [33]). By this interpretation of quantum mechanics it does not seem allowed to assume the possibility of attributing the result of a quantum measurement to the object as property already possessed *before* the measurement. Here quantum observables are attributed as properties to

quantum mechanical objects. Thus a property of the quantum system is defined only in the *context* of a measurement. There is no difference in the mathematical description between f-realism and contextualistic realism. In both frameworks an observable can be described as a function $U = U(\omega, \lambda)$, where ω and λ are hidden states of a measurement device and quantum system, respectively.

8. Other probabilistic models which do not contradict to local realism. L. Accardi [1] used non-Kolmogorovean model without Bayes' formula to eliminate Bell's inequality from considerations related to spin's model. Recently he also developed a new non-Kolmogorovean model which gives an explanation of violations of Bell's inequality, see [2].

I. Pitowsky [92], [93] discussed the possibility that some nonmeasurable sets can be physical events, i.e, some physical observables may be nonmeasurable. There is no Bell's inequality in this approach. Thus there is no problem with violations of Bell's inequality. This model is consistent with known polarization phenomena and the existence of macroscopic magnetism. He also proposed a thought experiment which indicates a deviation from the predictions of quantum mechanics. We noted that already A. N. Kolmogorov discussed 'generalized probabilities' on the algebra F_Ω of all subsets of Ω. Pitowsky discussed the relation of 'Banach-Tarski paradox' (Theorem 5.1, Chapter 1) to foundations of probability theory. It seems that Kolmogorov suspected that 'nonmeasurable events' could play some role in probability theory. The model of Pitowsky gives the interesting application of such 'generalized probabilities.' I. Pitowsky noticed:

"This so called 'Banach-Tarski paradox' is not a paradox at all. The pieces into which the ball is cut are nonmeasurable sets, that is, one cannot assign them numbers that indicate their volume since this will clearly violate the additivity or invariance of 'volume'. In spite of this explanation and in spite of independent proofs that nonmeasurable sets exist, the Banach-Tarski result was taken as an unfortunate consequence of the axiom of choice (which is nevertheless, essential in some fields of 'good' mathematics). Suppose, however, that we reverse this attitude and maintain that the subsets into which the ball is decomposed *exist in physical reality.* These hidden pieces could be detected in two 'states'. The first is a 'one-ball state' and the second a 'two-ball state'. In each state the pieces do have a 'volume' which depends, however, on their mutual configuration. Assume that we have a source that emits five balls in the first state. On the way from the source to a counter two of the balls spontaneously transform to the second state. The counter, which does not distinguish between the states, will detect seven ball. This rather simplistic example serves to indicate that one can 'perform miracles' if

one is willing to accept the physical reality of some highly abstract set-theoretical objects. In particular, if such assumptions are made, it is possible to account for interference effects in a completely mechanistic way without introducing wavelike nonlocal components to the theory.

Mathematicians, in particular applied mathematicians, where reluctant to take nonmeasurable sets seriously. As a result there exists no mathematical theory that relates nonmeasurable distributions with relative frequencies."

Such an extension of probability theory was created by I. Pitowsky and then strongly mathematically improved by S.P. Gudder [42]. He introduced the concept of a probability manifold M. The global properties of M inherited from its local structure were then considered. It was shown that a deterministic spin model due to Pitowski falls within this general framework. Finally, Gudder constructed a phase-space model for nonrelativistic quantum mechanics. These two models give the same global description as conventional quantum mechanics. However, they also give a local descriptions which is not possible in conventional quantum mechanics.

Remark 9.1. Non-Kolmogorovean probabilistic models of Accardi, Pitowski and Gudder have higher level of abstraction than the original Kolmogorov model. This is one of explanations why these models are not so popular in quantum physics. On the other hand, we showed that Bell's inequality does not contradict to local realism on the basis of the primary (rather primitive from mathematical viewpoint) probabilistic models, namely, the ensemble ans frequency models. It seems that our models have more close relation to physical reality.

We shall discuss in Chapter 3 the use of negative probabilities and in Chapter 4 the use of p-adic probabilities to eliminate Bell's inequality from considerations.

Conclusion. *'Bell's inequality' does not imply nonexistence of local hidden variables. Physical reality may be nonlocal. Physical reality may be nonobjective. However, both these features of physical reality are not related in any way to Bell's inequality.*

Chapter 3

Negative probabilities

In this chapter we study possibilities to extend the probability theory to describe numerous physical models with negative probabilities. Of course, negative probabilities could not appear in Kolmogorov's probability theory in that the probabilities of events must be **positive real numbers.** Therefore we have to turn back to the original probability formalisms, namely, ensemble and frequency.

1 The origin of negative probabilities in the ensemble and frequency theories

1. Ensemble approach: fluctuations of finite approximations. In the ensemble framework negative probabilities could not appear for finite statistical ensembles $S_N = \{s_1, s_2, ..., s_N\}$. However, such generalized probabilities can naturally appear for infinite statistical ensembles S as the results of the limit procedure:

$$\mathbf{P}_S(A = \alpha) = \lim_{N \to \infty} \frac{|S(A = \alpha) \cap S_N|}{|S|}, \tag{1.1}$$

where a sequence of finite ensembles $\{S_N\}$ gives an approximation of the infinite ensemble S. If this limit does not exist in $\mathbf{R}$, then some regularization procedures (for example, the summation of divergent series or integrals) can induce negative values for $\mathbf{P}_S(A = \alpha)$. Of course, in such a situation it would be natural to leave the domain of real analysis and consider some non-Archimedean number systems which contain actual infinities. In this case

the probability $\mathbf{P}_S(A = \alpha)$ can be defined directly as the proportion:

$$\mathbf{P}_S(A = \alpha) = \frac{|S(A = \alpha) \cap S_N|}{|S|} \,. \tag{1.2}$$

In Chapter 4 we shall use the system of p-adic numbers $\mathbf{Q}_p$ for such a purpose (another natural possibility is to use nonstandard numbers, [3]). In $\mathbf{Q}_p$ proportion (1.2) can be a negative rational number (as well as a rational number which is larger than 1).

2. Ensemble approach: split of conventional probabilities. Nonexistence of limits (1.1) is not the unique source of negative probabilities for infinite ensembles S. It may be situations (see, for example, the p-adic framework) such that limit (1.1) (for some α) exists and equal to zero (from the viewpoint of the real analysis). For example, for the uniform distribution on $S = \mathbf{N}$, we have $\mathbf{P}_S(A = n) = \lim_{N\to\infty} \frac{1}{N} = 0$ for all $n = 1, 2, \ldots$. However, some regularization of this limit procedure can produce nonzero coefficients $\mathbf{P}_S^{\rm reg}(\alpha)$. In the mentioned p-adic framework such coefficients (defined by (1.1) with respect to the p-adic topology or directly by (1.2) with the aid of actual infinities) are always negative (rational) numbers [1]. Thus regularizations of (1.1) can induce the split of zero conventional probabilities in a set of new labels which can be negative numbers. These new labels can be interpreted as infinitely small probabilities. Such a split of conventional probabilities is not a feature of only zero probabilities. For example, probability one can be also split in a set of new labels which are interpreted as probabilities which differ from probability one by infinitely small probabilities. These are 'practically one probabilities'. In all p-adic examples such new probabilities are given by rational numbers which are larger than one[2]. Similar splits can be obtained for other rational probabilities $q \in (0, 1)$. If $0 < q < 1, q \in \mathbf{Q}$, then we have two sets of labels $L_{<q}$ and $L_{>q}$. They denote, respectively, probabilities $a = q - \lambda$ and $a = q + \lambda$, where λ is infinitely small probability. In p-adic examples we have $L_{<q} \subset \mathbf{Q} \cap (-\infty, 0)$ and $L_{>q} \subset \mathbf{Q} \cap (1, +\infty)$ (see Chapter 4 for the details).

On one hand, probabilities $q < 0$ (and $q > 1$) demonstrate irregular behaviour ($N \to \infty$) of approximations of probabilities $\mathbf{P}_S$ with respect to an

[1]In fact, we could not prove such a general theorem in the framework of p-adic analysis. But numerous examples demonstrate this feature of the p-adic split of zero conventional probabilities.

[2]This is natural: if $\mathbf{P}(A) = q < 0$ is infinitely small probability, then $\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A) = 1 - q > 1$ is probability which negligibly differs from 1 and vice versa.

infinite ensemble S by probabilities $\mathbf{P}_{S_N}$with respect to finite sub-ensembles S_N. On the other hand, they can describe the fine internal structure of S (via split of conventional probabilities). We note that from the physical point of view the irregularity of approximations means that it is impossible to prepare for all measurements for a quantum state ϕ (describing S) finite ensembles S_N with identical statistical properties.

3. Frequency approach: irregularity of behaviour of frequencies. In the frequency framework negative probabilities could not appear in the classical theory of R. von Mises which is based on the principle of the statistical stabilization of frequencies with respect to the real metric. However, if we assume that for some 'quasi-random sequences' $x = (x_1, x_2, ..., x_n, ...)$ this principle can be violated, namely, the limit

$$\mathbf{P}_x(\alpha) = \lim_{N \to \infty} \nu_N(\alpha; x) \tag{1.3}$$

does not exist in $\mathbf{R}$, then some regularization procedures $\mathcal{R}_{\mathrm{fr}}$ for (1.3) can produce negative values (as well as values which are larger than 1) for $\mathbf{P}_x$. One of the possibilities for such a regularization is to change the topology on the set of rational numbers $\mathbf{Q}$ in that we study the convergence of relative frequencies. In Chapter 4 we shall use the p-adic topology for such a purpose.

4. Frequency approach: split of Mises' probabilities. Another source of frequency probabilities $q < 0$ and $q > 1$ is the split of Mises' probabilities. For example, the fact that frequency probability $\mathbf{P}_x^{\mathrm{Mises}}(A) = \lim_{n \to \infty} \nu_n(A; x) = 0$ does not imply that the event A should never occur. Therefore it is reasonable to take such events into account by using new labels.

Let us consider two events A and B which have zero frequency probabilities:

$$\mathbf{P}_x^{\mathrm{Mises}}(A) = \lim_{n \to \infty} \nu_n(A; x) = 0, \ \mathbf{P}_x^{\mathrm{Mises}}(B) = \lim_{n \to \infty} \nu_n(B; x) = 0, \tag{1.4}$$

in $\mathbf{R}$. We are interested in the problem: What event, A or B, has larger probability? Of course, this question is meaningless from the viewpoint of the Mises' probability theory. However, this problem can be solved by extending the set of labels for probabilities.

In the frequency framework we can obtain new sets of labels automatically by using new topologies for the statistical stabilization (by finding limits (1.3)

with respect to new topologies)[3]. Each topology of the statistical stabilization induces its own set of labels for split Mises' probabilities. For example, it may be that, despite of (1.4) in $\mathbf{R}$, we have

$$\mathbf{P}_x^\tau(A) = \lim_{n\to\infty} \nu_n(A;x) \neq 0,\ \mathbf{P}_x^\tau(B) = \lim_{n\to\infty} \nu_n(B;x) \neq 0 \qquad (1.5)$$

for some topology τ on $\mathbf{Q}$. If we choose the p-adic topology $\tau = \tau_p$, then in examples studied by the author p-adic probabilities (1.5) are represented by negative rational numbers. Thus by using negative probabilities we can split zero (Mises') probability. The same split can be obtained for all Mises' probabilities $q \in [0,1] \cap \mathbf{Q}$.

On one hand, probabilities $q < 0$ and $q > 1$ demonstrate the violation of the principle of the statistical stabilization (the law of large numbers) for some 'quazi-random' sequences. On the other hand, they describe (with the aid of new topologies on $\mathbf{Q}$) the fine internal structure of some Mises' collectives.

5. Where are negative probabilities? However, the reader may ask: Why could we not find negative probabilities in physical experiments? One of reasons is that, in fact, we have never tried to find them. All our experimental methodology is based on the principle of the statistical stabilization (the law of large numbers). All experiments are prepared in such circumstances that relative frequencies must stabilize. This is the result of our cognitive evolution. In the process of evolution the brain extracted from the chaotic and (lawless) reality phenomena which satisfy the principle of the statistical stabilization (repeatability in the average). These and only these phenomena are considered by the brain as real physical phenomena. Negative probabilities give the possibility to extend the range of physical phenomena by considering phenomena which violate the principle of the statistical stabilization. Another reason of the absence of negative probabilities in the experimental framework is the common use of real analysis for the study of the experimental statistical data. However, this data is always rational and in principle other topologies on $\mathbf{Q}$ (different from the real one) can be used for studying of this data. In particular, we have to pay more attention to events A with zero conventional (Kolmogorov or Mises) probabilities, $\mathbf{P}^{\mathrm{Conv}}(A) = 0$. From our viewpoint such events are not less physical than events with positive probabilities. By using negative probabilities we can consider in analytical calculations events A such that $\mathbf{P}^{\mathrm{Conv}}(A) = 0$. In this way we can clarify the hidden internal structure

[3]The real topology is only one of many topologies on the set of rational numbers $\mathbf{Q}$ which contains frequencies $\nu_N = n/N$.

of some events B with positive conventional probabilities. We shall study this question carefully in the next section.

6. The formula of total probability as an average procedure. We consider a quantum measurement for quantum systems prepared in a state ϕ. We suppose that each quantum system s which is taken for this measurement has a hidden state λ which determines (with some probability) a result of the measurement for the s (see Chapter 2). The set of hidden states is denoted by Λ. The number of hidden states may be infinite.

Remark 1.1. Of course, in a laboratory we can produce only a finite ensemble $S_N = \{s_1, ..., s_N\}$ of quantum systems which have a finite number of hidden states $\lambda_1, ..., \lambda_n, n \leq N$. However, different finite ensembles $S_N, \tilde{S}_M, ...$ are used in different experiments. It is natural to assume that these finite ensembles are subensembles of one infinite ensemble S. The quantum state ϕ describes this infinite ensemble. The infinite cardinality of S induces the impression that S is just an ideal mathematical abstraction. However, suppose, for example, that each electron s has the extremely complex internal structure. Then, in fact, each s must be described by its own (individual) internal state λ. In this case the number of all possible states (for all electrons in the universe) is really infinite.

The (hidden) probability for λ in S is denoted by the symbol $\mathbf{p}_\lambda$. On the basis of our previous considerations (see Chapter 2) it is natural to suppose that some of $\mathbf{p}_\lambda$ may be nonconventional probabilities; in particular, they may be negative[4].

In the process of a measurement each state λ is transformed into a new state λ' (due to an interaction between the quantum system and the equipment). Denote probabilities of this transition by $\mathbf{p}_{\lambda\lambda'}$. Some of these probabilities can be negative (in particular, the law of large numbers can be violated for some transitions $\lambda \to \lambda'$). In the measurement we observe events A consisting of some sets of states λ' (in principle these sets can be infinite). By

[4]In particular, they may be infinitely small probabilities. For example, if each electron in the universe has its own state λ, then $\mathbf{p}_\lambda = \lim \frac{1}{N} = 0$ (from the viewpoint of real analysis). Negativity of $\mathbf{p}_\lambda$ can also be a consequence of the violation of the law of large numbers. Such a violation for hidden states λ is quite natural if $|\Lambda| = \infty$. For the concrete λ, behaviour of frequencies $\nu_N(\lambda; x)$ can strongly depend on a sample x. There are no reasons to assume that two different samples of quantum systems $S_N = \{s_1, ..., s_N\}$ and $\tilde{S}_M = \{\tilde{s}_1, ..., \tilde{s}_M\}$ must produce samples $x = (\lambda_1, ..., \lambda_N)$ and $\tilde{x} = (\tilde{\lambda}_1, ..., \tilde{\lambda}_M)$ having the same probability distribution (because our macro equipment could not control statistical behaviour of hidden parameters).

the formula of total probability we obtain:

$$\mathbf{P}(A) = \sum_{\lambda \in \Lambda} \mathbf{p}_\lambda \sum_{\lambda' \in A} \mathbf{p}_{\lambda\lambda'} . \qquad (1.6)$$

In fact, this is the average procedure with respect to the ensemble Λ of hidden states λ, transitions $\lambda \to \lambda'$ and states λ' which are identified in the observed event A. The result of this procedure can be a conventional (Kolmogorov or Mises) probability, despite of the possibility that some of probabilities $\mathbf{p}_\lambda, \mathbf{p}_{\lambda\lambda'} < 0$ or $\mathbf{p}_\lambda, \mathbf{p}_{\lambda\lambda'} > 1$.

The ensemble and frequency explanations of this phenomenon have been already presented in section 4, Chapter 2. For example, in the frequency framework fluctuations of frequencies $\nu_N(\lambda)$ and (or) $\nu_N(\lambda'/\lambda)$ can compensate each other and produce the statistical stabilization. Examples 4.1 and 4.2 showed that such a behaviour can be demonstrated even in the case of a finite set Λ. Thus one of the sources of conventional probabilities in (1.6) is that simultaneous (chaotic) fluctuations can produce in average the statistical stabilization. Another source are infinite statistical ensembles with infinitely small initial probabilities $\mathbf{p}_\lambda < 0$ and (or) transition probabilities $\mathbf{p}_{\lambda\lambda'} < 0$. Infinite sums of infinitely small (negative) probabilities might produce conventional positive probabilities.

In all previous considerations the formula of total probability must be regularized via some procedure (for example, by using a new number system to find the limits of fluctuating frequencies, see Chapter 4). In general we could not even suppose the validity of the Bayes' formula (even for one fixed state λ and transition $\lambda \to \lambda'$).

Example 1.1. Example 4.1 (Chapter 2) can be generalized by considering the infinite set of hidden states $\lambda \in [0, \pi]$. We choose the uniform probability distribution on $[0, \pi]$ as the initial probability distribution $\mathbf{p}_\lambda$ (these are infinitely small probabilities). However, in the framework of real analysis we could not represent $\mathbf{p}_\lambda$ as proportional probabilities (1.2). The only thing which we can do is to use normalized Lebesgue measure on $[0, \pi]$ to represent $\mathbf{p}_\lambda$. Let us consider an observable $B = 0, 1$ ($\lambda' = B$) with conditional frequencies

$$\nu_k(0/\lambda) \approx \sin^2 k\lambda,\ \nu_k(1/\lambda) \approx \cos^2 k\lambda,\ k \to \infty, \lambda \in [0, \pi] .$$

If $\lambda \neq \pi l, l = 0, 1, 2, ...$, then conditional frequency probabilities $\mathbf{P}^{\text{fr}}(B = 0/\lambda) = \lim_{k\to\infty} \sin^2 k\lambda$ and $\mathbf{P}^{\text{fr}}(B = 1/\lambda) = \lim_{k\to\infty} \cos^2 k\lambda$ do not exist.

But the average procedure based on the (integral) formula of total probability gives well defined conventional probabilities for values of B :

$$\mathbf{P}(B=0)=\lim_{k\to\infty}\int_0^{\pi}\sin^2 k\lambda\, d\mathbf{p}_\lambda=\frac{1}{2},\ \mathbf{P}(B=1)=\lim_{k\to\infty}\int_0^{\pi}\cos^2 k\lambda\, d\mathbf{p}_\lambda=\frac{1}{2}.$$

In section 3 we shall study examples in that nonexistence of conventional conditional probabilities implies negativity of generalized conditional probabilities.

Example 1.2. The previous example can be easily modified to obtain a model in that probabilities $\mathbf{P}^{\mathrm{fr}}(\lambda)=\mathbf{p}_\lambda$ do not exist. Let $d\nu_k(\lambda)\approx\frac{2}{\pi}\sin^2 k\lambda\ d\lambda, k\to\infty$, and let $\nu_k(B=\beta/\lambda)\approx\frac{1}{2}, k\to\infty$. Then the frequency probability distribution $\mathbf{p}_\lambda$ do not exist. But via the formula of total probability we obtain in the average:

$$\mathbf{P}(B=0)=\lim_{k\to\infty}\frac{1}{2}\int_0^{\pi}d\nu_k(\lambda)=\frac{1}{2}\ ,\mathbf{P}(B=1)=\lim_{k\to\infty}\frac{1}{2}\int_0^{\pi}d\nu_k(\lambda)=\frac{1}{2}.$$

7. Negative probabilities and the principle of complementarity. The considerations of the previous section on the formula of total probability as an average procedure are based on ideas of P. Dirac [34] and R. Feynman [37]. In particular, R. Feynman considered a roulette which has two internal (non-observed) states λ_1 and λ_2 and three observed states 1,2,3. By simple numerical examples (that the reader can produce by himself) he demonstrated that observed events can have positive conventional probabilities $\mathbf{p}_j>0, j=1,2,3$, despite of negativity of some hidden probabilities $\mathbf{p}_{\lambda_1},\mathbf{p}_{\lambda_2}$ or conditional probabilities $\mathbf{p}_{\lambda_1 j},\mathbf{p}_{\lambda_2 j}, j=1,2,3$. However, neither Dirac nor Feynman could propose a mathematical explanation of the origin of negative probabilities (they considered negative probabilities as just formal quantities which could be useful in some calculations). I have found the frequency and ensemble roots of negative probabilities. For example, we can build Feynman's roulette by using 'quasi-random' generators for states λ_1 and λ_2 or for transitions $\lambda_1\to j$ and $\lambda_2\to j$ which simulate the statistical models of Examples 4.1 and 4.2 (Chapter 2), respectively.

On the basis of our interpretation of negative probabilities it would be interesting to discuss the idea of R. Feynman on a connection between negative probabilities and the principle of complementarity in quantum mechanics, see [37]. As I could understood, R. Feynman is an adherent of i-realism (at least in this paper).

In the framework of subsection 6 we consider two physical properties A and B. Thus (despite of possible fluctuations of frequencies and conditional frequencies for hidden variables) frequencies

$$\nu_N(A_\alpha) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha} \nu_N(\lambda'/\lambda) \text{ , where } A_\alpha = \{A = \alpha\} \text{ ,} \tag{1.7}$$

$$\nu_N(B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in B_\beta} \nu_N(\lambda'/\lambda) \text{ , where } B_\beta = \{B = \beta\} \text{ ,} \tag{1.8}$$

stabilize (when $N \to \infty$) to conventional probabilities $\mathbf{P}^{\mathrm{Conv}}(A_\alpha), \mathbf{P}^{\mathrm{Conv}}(B_\beta)$.

However, in general there are no reasons to suppose that the frequency

$$\nu_N(A_\alpha \cap B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha\cap B_\beta} \nu_N(\lambda'/\lambda) \tag{1.9}$$

also stabilize (when $N \to \infty$)[5]. If (1.9) does not stabilize, then conventional probability $\mathbf{P}^{\mathrm{Conv}}(A = \alpha, B = \beta)$ is not defined.

Remark 1.2. Suppose that we could find some procedure $\mathcal{R}_{\mathrm{fr}}$ to regularize fluctuating frequencies $\nu_N(\lambda)$ and (or) $\nu_N(\lambda'/\lambda)$. By $\mathcal{R}_{\mathrm{fr}}$ we obtain generalized probabilities $\mathbf{p}_\lambda$ and (or) $\mathbf{p}_{\lambda\lambda'}$ (which in principle can be negative numbers). Suppose that (in the case of the infinite set Λ) we could find some procedure $\mathcal{R}_{\mathrm{conv}}$ to regularize (probably diverging) series

$$\sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in A_\alpha} \mathbf{p}_{\lambda\lambda'} \, , \sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in B_\beta} \mathbf{p}_{\lambda\lambda'} \tag{1.10}$$

in such a way that their sums coincide with conventional probabilities $\mathbf{P}^{\mathrm{Conv}}(A_\alpha)$ and $\mathbf{P}^{\mathrm{Conv}}(B_\beta)$, respectively. We now apply $\mathcal{R}_{\mathrm{conv}}$ to series

$$\sum_{\lambda\in\Lambda} \mathbf{p}_\lambda \sum_{\lambda'\in A_\alpha\cap B_\beta} \mathbf{p}_{\lambda\lambda'} \, . \tag{1.11}$$

In principle there may be different variants: 1) the procedure $\mathcal{R}_{\mathrm{conv}}$ does not work for series (1.11); here we could not assign any real number to (1.11); 2) despite of fluctuations of frequencies (1.9), the $\mathcal{R}_{\mathrm{conv}}$ still works for series (1.11) and gives a real number; but this number is not related to the statistical limit of frequencies (1.9) (in particular, it may be a negative number).

[5]If $|\Lambda| = \infty$, then events A_α and B_β may differ rather slightly: $\nu_N(A/\lambda) \approx \nu_N(B/\lambda)$ for each $\lambda \in \Lambda$. But the infinite average over Λ can produce behaviour of frequencies (1.9) which essentially differs from behaviour of frequencies (1.7) and (1.8).

This simple statistical consideration explains the origin of difficulties with 'simultaneous existence' of incompatible properties of quantum systems. Therefore the presence of incompatible properties does not demonstrate some essentially new 'quantum' properties of reality. It only demonstrates that the law of large numbers is violated for internal (hidden) properties of so called quantum systems (mainly because we could not control the statistical behaviour of these properties in our (macro) preparation procedures). For some events fluctuations on the microlevel can compensate each other and produce the statistical stabilization of observed frequencies (1.7) and (1.8). At the present time such events are called physical events. For other events fluctuations on the microlevel cannot compensate each other; there is no statistical stabilization of observed frequencies (1.9). At the present time such events are called nonphysical.

There are also no reasons to suppose that (in general generalized) initial probability distribution $\mathbf{p}_\lambda$ and conditional probabilities $\mathbf{p}_{\lambda\lambda'}$ can be chosen in such a way that fluctuations in both expressions (1.7) and (1.8) could be compensated so that, for some values $A = \alpha_0$ and $B = \beta_0$, both frequencies $\nu_N(A_{\alpha_0})$ and $\nu_N(B_{\beta_0})$ stabilize to probability 1. This is nothing than the statistical explanation of the principle of the complementarity. It seems that (rather unclear) considerations of R. Feynman [37] can be interpreted in such a way.

Thus we proposed the purely statistical explanation of the phenomenon of incompatibility for some quantum observables. Here the problem of disturbance effects of measurements is totally excluded from considerations. Our approach implies that even the possibility to perform measurements on quantum systems without any disturbance effect would not imply that incompatible properties can be measured simultaneously[6]. Different structure of sets $\{\lambda' \in A_\alpha\}$, $\{\lambda' \in B_\beta\}$ and $\{\lambda' \in A_\alpha \cap B_\beta\}$ might still imply fluctuations of frequencies (1.9).

Thus the careful probabilistic considerations show that there may exist physical (in the sense of the verification by the law of large numbers) properties A, B such that the simultaneous existence of these properties could not be verified on the physical level. In such a situation one of the possibilities is to exclude pairs

[6]The idea that the presence of incompatible observables in the quantum formalism (and, in particular, the Heisenberg uncertainty relation) is not a consequence of disturbance effects of the process of a measurement, but a consequence of the internal statistical structure of a quantum state (or a preparation procedure), has been intensively discussed in quantum physics (Prugovecki [94], Ballentine [8]).

$C = (A, B)$ of incompatible properties from considerations (this is the modern quantum viewpoint) [7]. However, there is another possibility, namely, consider some regularization procedure $\mathcal{R}$ for (1.9). If (1.9) could be regularized via $\mathcal{R}$, then C can be considered as $\mathcal{R}$-physical property. Thus we can essentially extend physical reality by considering new $\mathcal{R}$-elements of reality. As we have already remarked, in many cases one of the simplest ways to regularize (1.9) is to use the p-adic topology, instead of the real. Here frequencies $\nu_N(A_\alpha \cap B_\beta)$ may have the limit in $\mathbf{Q}_p$, despite of fluctuations in $\mathbf{R}$. However, the possibility of a p-adic (and any other) regularization of (1.9) need not imply the possibility to use the same regularization for (1.7) and (1.8). In principle A and B need not be elements of new reality (despite of the fact that $C = (A, B)$ is an element of this reality). Nevertheless, there may be coincidences such that all series

$$\mathbf{P}^{\mathcal{R}}(A_\alpha) = \sum_{\lambda \in \Lambda} \mathbf{p}_\lambda \sum_{\lambda' \in A_\alpha} \mathbf{p}_{\lambda\lambda'} \,, \mathbf{P}^{\mathcal{R}}(B_\beta) = \sum_{\lambda \in \Lambda} \mathbf{p}_\lambda \sum_{\lambda' \in B_\beta} \mathbf{p}_{\lambda\lambda'} \,. \tag{1.12}$$

$$\mathbf{P}^{\mathcal{R}}(A_\alpha \cap B_\beta) = \sum_{\lambda \in \Lambda} \mathbf{p}_\lambda \sum_{\lambda' \in A_\alpha \cap B_\beta} \mathbf{p}_{\lambda\lambda'} \tag{1.13}$$

converge with respect to $\mathcal{R}$. In such a situation all events, $A_\alpha, B_\beta, A_\alpha \cap B_\beta$ are $\mathcal{R}$-physical events. It could be that $\mathcal{R}$-probabilities $\mathbf{P}^{\mathcal{R}}(A_\alpha)$ and $\mathbf{P}^{\mathcal{R}}(B_\beta)$ coincide with conventional probabilities $\mathbf{P}^{\mathrm{Conv}}(A_\alpha)$ and $\mathbf{P}^{\mathrm{Conv}}(B_\beta)$. However, in general $\mathbf{P}^{\mathcal{R}}(A_\alpha) \neq \mathbf{P}^{\mathrm{Conv}}(A_\alpha)$, and (or) $\mathbf{P}^{\mathcal{R}}(B_\beta) \neq \mathbf{P}^{\mathrm{Conv}}(B_\beta)$. In the ensemble framework the previous considerations can be interpreted in the following way. The system of events $F(\pi_S)$ for the ensemble S need not be an algebra. The sets $C_{\alpha\beta} = A_\alpha \cap B_\beta$ need not belong to $F(\pi_S)$. However, we may try to extend the ensemble probability to larger class of sets by using some regularization procedures. Sometimes it is possible and sometimes it is impossible to define ensemble probabilities for $C_{\alpha\beta}$ and preserve ensemble probabilities for sets A_α and B_β.

Thus the modern physics is based the *Kolmogorov physical reality.* This model of physical reality can be extended by considering *non-Kolmogorov physical realities.* We conclude our considerations by the equality:

Model of Reality = Model of Probability.

We now consider the principle of complementarity in the framework of f-realism. The main difference between i-realism and f-realism is that in the

[7] E. Prugovecki pointed out [94] that, far from restricting simultaneous measurements of noncommuting observables, quantum theory does not deal with them at all; its formalism being capable only of statistically predicting the results of measurements of one observable (or a commuting set of observables).

first case we can assume that conditional probabilities $\mathbf{p}_{\lambda\lambda'}$ do not depend on a measured property and in the second case a measurement of a property D produces $\mathbf{p}_{\lambda\lambda'} = \mathbf{p}^D_{\lambda\lambda'}$. Here

$$\nu_N(A_\alpha) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha} \nu_N(\lambda'/\lambda; A) \ , \tag{1.14}$$

$$\nu_N(B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in B_\beta} \nu_N(\lambda'/\lambda; B) \tag{1.15}$$

$$\nu_N(A_\alpha \cap B_\beta) = \sum_{\lambda\in\Lambda} \nu_N(\lambda) \sum_{\lambda'\in A_\alpha\cap B_\beta} \nu_N(\lambda'/\lambda; C), \ C = (A, B) \ . \tag{1.16}$$

Here (even for finite sets Λ of hidden variables) the statistical stabilization of frequencies (1.14) and (1.15) need not imply the statistical stabilization of frequencies (1.16).

8. History of negative probabilities in physics. The possibility to obtain negative probabilities via a regularization of ensemble and frequency approximations (1.1) and (1.3), respectively, is so natural that the negative attitude against negative probabilities in physics can be only explained by the common use of Kolmogorov' theory of probability. From the frequency viewpoint this use imply the common viewpoint that relative frequencies must always stabilize; from the ensemble viewpoint this use imply that statistical ensembles of physical systems must always have a homogeneous structure with respect to all their nonobserved properties. A negative psychological reaction to the appearance of negative probabilities in physical models implies the desire to forget papers in that negative probabilities play the fundamental role.

Although it is well known, for instance, that P. A. M. Dirac was the first to introduce explicitly the concept of negative energy, the number of those who know his investigations [34] about negative probability - closely related to negative energy and invented simultaneously-seems to be very restricted. This concept is used with reservations but, as it seems, not without a certain kind of sympathy. Said paper (see section ? for the details) is not the only one on this topic meanwhile has been forgotten, at least as far as negative probability is concerned.

Another example is the famous Wiegner distribution [111] $W(q, p)$ which had been introduced as a probability distribution (see section 4). And it has no other physical interpretation than a probability distribution. However, the appearance of negative probabilities for some quantum states implies that Wiegner's distribution is not more interpreted as a probability distribution (many physicists prefer to call $W(q, p)$ Wiegner's function).

In the framework of the EPR experiments violations of Bell's inequality could be easily explained if we suppose that there exist negative probability distributions. However, papers on negative probability description of the EPR experiments (see, for example, the review of W. Muckenheim [90]) did not play large role in the polemics on the EPR experiments. Physicists prefer to accept the death of reality (namely, the impossibility to use realism in quantum world; thus, in fact, the absence of objective laws in reality) or nonlocality of space-time than to use negative probabilities.

The existence of quantum observables with continuous spectra is in the evident contradiction with the discreteness of results of real physical measurements. E. Prugovecki [94] developed a theory of quantum measurements with a finite precision (which takes into account reading errors of individual measurements). One of the great advantages of this theory is the possibility to describe simultaneous measurements of incompatible observables. However, there appear again negative probabilities[8]. As always, this implied the extremely strong critic of the theory.

2 Signed 'probabilistic' measures and Einstein-Podolsky-Rosen paradox

We start this section with brief mathematical introduction to the theory of signed measures (charges). Let Ω be a set and let $\mathcal{F}$ be a σ-algebra of its subsets. A σ-additive function $\mu : \mathcal{F} \to \mathbf{R}$ is said to be a *signed measure* (*charge*). Thus $\mu(\cup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} \mu(A_n)$ for any sequence $A_n \in \mathcal{F}$, $A_j \cap A_i = \emptyset, i \neq j$.

Example 2.1. (Discrete measures) Let $\Omega = \{x_1, x_2, ..., x_n, ...\}$ be a countable set and let $\mathcal{F}$ be a σ-algebra of all subsets of Ω. Let $\{a_n\}_{n=1}^{\infty}$ be a sequence of real numbers such that $\sum_{n=1}^{\infty} |a_n| < \infty$. We set $\mu(\{x_n\}) = a_n$ and $\mu(A) = \sum_{x_n \in A} \mu(\{x_n\})$ for $A \in \mathcal{F}$. The $\mu : \mathcal{F} \to \mathbf{R}$ is a signed measure. On the basis of this simple example we illustrate some important notions of the general theory of signed measures. Set $\Omega_- = \{x_j \in \Omega : \mu(\{x_j\}) < 0\}$, $\Omega_+ = \{x_j \in \Omega : \mu(\{x_j\}) > 0\}$ and $\Omega_0 = \{x_j \in \Omega : \mu(\{x_j\}) = 0\}$. It is evident that for any $E \in \mathcal{F}$:

$$\mu(E \cap \Omega_-) \leq 0 \quad \text{and} \quad \mu(E \cap \Omega_+) \geq 0. \tag{2.1}$$

[8]This has the natural explanation on the basis of our interpretation of negative probabilities: the violation of the law of large numbers for such measurements.

Let $U, V \in \mathcal{F}, U \cap V = \emptyset$ and let $\Omega_0 = U \cup V$. Set $\Omega'_- = \Omega_- \cup U$ and $\Omega'_+ = \Omega_+ \cup V$ (thus $\Omega = \Omega'_- \cup \Omega'_+$). Then the sets Ω'_- and Ω'_+ has same property (2.1) as the sets Ω_- and Ω_+. Set

$$\mu^-(E) = -\mu(E \cap \Omega'_-) = \sum_{x_n \in E \cap \Omega'_-} |a_n| \text{ and } \mu^+(E) = \mu(E \cap \Omega'_+) = \sum_{x_n \in E \cap \Omega'_+} a_n.$$

Then $\mu(E) = \mu^+(E) - \mu^-(E)$. This representation of μ is unique (in spite of nonuniqueness of a representation $\Omega = \Omega'_- \cup \Omega'_+$). We can associate with a signed measure μ the positive measure $|\mu| = \mu^+ + \mu^-$, $\mu(A) = \sum_{x_n \in A} |a_n|$.

In fact, this particular example demonstrated all main features of signed measures. We consider now the general case.

Definition 2.1. *Let μ be a signed measure defined on a σ-algebra $\mathcal{F}$ of subsets of a space Ω. Then the set $A \subset \Omega$ is said to be* **negative** *with respect to μ if $E \cap A \in \mathcal{F}$ and $\mu(E \cap A) \leq 0$ for every $E \in \mathcal{F}$. Similarly, A is said to be* **positive** *with respect to μ if $E \cap A \in \mathcal{F}$ and $\mu(E \cap A) \geq 0$ for every $E \in \mathcal{F}$.*

Theorem 2.1. (Hahn-Jordan) *Given a signed measure μ on a σ-algebra $\mathcal{F}$ of subsets of Ω, there exists a set $\Omega_- \in \mathcal{F}$ such that Ω_- is negative and $\Omega_+ = \Omega \setminus \Omega_-$ is positive with respect to μ.*

Proof. Let $a = \inf \mu(A)$ where the greatest lower bound is taken over all negative sets $A \in \mathcal{F}$. Let $A_n \in \mathcal{F}, n = 1, 2, ...$, be a sequence of negative sets such that $\lim_{n\to\infty} \mu(A_n) = a$. Then the set $\Omega_- = \cup_{n=1}^{\infty} A_n \in \mathcal{F}$ is a negative set such that $\mu(\Omega_-) = a$ (this is a consequence of σ-additivity of μ). To show that Ω_- is the required set, we must only show that $\Omega_+ = \Omega \setminus \Omega_-$ is positive. It is possible to show that the assumption Ω_+ is not positive will imply the contradiction (see, for example, [78] for the details). ■

Thus we can represent Ω as a union

$$\Omega = \Omega_+ \cup \Omega_- \tag{2.2}$$

of two disjoint measurable sets Ω_+ and Ω_-, where Ω_+ is positive and Ω_- is negative with respect to the signed measure μ. The representation (2.2) is called the *Hahn decomposition* of Ω, and may be not unique. However, if

$$\Omega = \Omega^1_+ \cup \Omega^1_-, \quad \Omega = \Omega^2_+ \cup \Omega^2_-$$

are two distinct Hahn decompositions of Ω, then

$$\mu(E \cap \Omega^1_-) = \mu(E \cap \Omega^2_-) \, , \ \mu(E \cap \Omega^1_+) = \mu(E \cap \Omega^2_+) \tag{2.3}$$

for every $E \in \mathcal{F}$. In fact, $E \cap (\Omega_-^1 \setminus \Omega_-^2) \subset E \cap \Omega_-^1$ and at the same time $E \cap (\Omega_-^1 \setminus \Omega_-^2) \subset E \cap \Omega_+^2$. This imply that

$$\mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) \leq 0 \text{ and } \quad \mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) \geq 0.$$

Thus $\mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) = 0$, and similarly $\mu(E \cap (\Omega_-^2 \setminus \Omega_-^1)) = 0$. Therefore

$$\mu(E \cap \Omega_-^1) = \mu(E \cap (\Omega_-^1 \setminus \Omega_-^2)) + \mu(E \cap (\Omega_-^1 \cap \Omega_-^2))$$

$$= \mu(E \cap (\Omega_-^2 \setminus \Omega_-^1)) + \mu(E \cap (\Omega_-^1 \cap \Omega_-^2)) = \mu(E \cap \Omega_-^2),$$

which proves the first of the formulas (2.3). The second formula is proved in exactly the same way.

Thus a signed measure μ on the space Ω uniquely determines two nonnegative set functions, namely

$$\mu^+(E) = \mu(E \cap \Omega_+), \quad \mu^-(E) = -\mu(E \cap \Omega_-)$$

called the *positive variation* and *negative variation* of μ, respectively. It is clear that

1) $\mu = \mu^+ - \mu^-$;

2) μ^+ and μ^- are nonnegative σ-additive set functions, i.e., measures;

3) The set function $|\mu| = \mu^+ + \mu^-$, called the *total variation* of μ, is also a measure.

The representation $\mu = \mu^+ - \mu^-$ is called the *Jordan decomposition* of μ.

We can present a formal generalization of Kolmogorov measure-theoretical approach. We define a *signed probability space* as the triple $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$, where Ω is an arbitrary set (points ω of Ω are said to be *elementary events*), $\mathcal{F}$ is an arbitrary σ-algebra of subsets of Ω (elements of $\mathcal{F}$ are said to be *events*), $\mathbf{P}$ is a σ-additive signed measure (a charge) on $\mathcal{F}$ normalized by the condition $\mathbf{P}(\Omega) = 1$.

It is a generalization of probability. There can be events which have negative probabilities and probabilities which are larger than 1. However, our consideration in subsection 1 give strong motivations to use signed probability spaces in physics. Moreover, there are analogues of the law of large numbers and central limit theorem for signed probabilities (see [9], [51], [52]) which also improve the use of signed probability spaces.

There are no physical reasons to assume that even in the case of signed probabilities the system of events has the structure of a set algebra (see also Chapter

4). It is natural to consider signed probability semi-measures defined on set semi-algebras (the reader can obtain the definition of signed probability semi-measure by analogue to Definition 5.2 of Chapter 1). However, it seems that the corresponding mathematical formalism is not yet developed. In particular, I do not know anything about the possibility to obtain the Jordan decomposition for signed semi-measures.

As it has been already mentioned, some physicists (see, for example, [90]) assume that probability distributions involved in Bell's considerations are signed probability measures. This assumption implies that we could not use the standard probabilistic estimates. Therefore there is **no Bell's inequality at all.** From this viewpoint experiments for testing Bell's inequality can be considered as **experiments for testing foundations of probability theory.**

We discuss now carefully the origin of negative probabilities in the EPR framework. Let us follow the ideology of hidden variables. Consider a number N of particles prepared in a pure quantum state and possessing hidden variables $\lambda_k, k = 0, ..., n$. Assume that the different values λ_k are taken with (probably generalized) probabilities $\mathbf{p}_k$. By an interaction (the nature of which need not be specified) the values of these hidden variables change from λ_i to $\lambda'_j, j = 0, ..., m$, the transition probability being denoted by $\mathbf{p}_{kj}$. By this interaction the pure state may split into $l \leq m$ experimentally distinguishable states. Let A be one of such states. The set of values of j such that λ'_j form the state A is denoted by the symbol $j(A)$. The result of a measurement exhibits $N(A)$ particles in the state A and gives relative frequencies $\nu_N(A) = \frac{N(A)}{N}$. By statistical stabilization of these frequencies we obtain frequency probabilities: $\mathbf{P}^{\mathrm{Mises}}(A) = \lim_{N\to\infty} \nu_N(A)$. The combined transition probability for the state A can be found with the aid of the formula of total probability:

$$\mathbf{P}^{\mathrm{com}}(A) = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbf{p}_k \sum_{j\in j(A)} \mathbf{p}_{kj} = \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{j\in j(A)} \mathbf{p}_{kj} \, . \tag{2.4}$$

All probabilistic considerations on Bell's inequality are based on the assumption that the observed frequency probabilities $\mathbf{P}^{\mathrm{Mises}}(A)$ must coincide with combined transition probabilities $\mathbf{P}^{\mathrm{com}}(A)$ (defined by (2.4)). By this assumption we can use hidden probabilities $\mathbf{p}_k, \mathbf{p}_{kj}$, in calculations related to Bell's inequality. However, as it has been already mentioned in section 1, the formula of total probability (2.4) can contain some pathologies. These

pathologies could be in principle eliminated by some regularization procedure $\mathcal{R}$ [9]. However, $\mathcal{R}$ can produce nonconventional probabilities $\mathbf{p}_j, \mathbf{p}_{ij}$.

The problem of fluctuating of frequencies $\nu_N(\lambda_k)$ and (or) $\nu_N(\lambda'_j/\lambda_k)$ have been already discussed in section 1 (see also Chapter 2). We pay now attention to the average over an infinite set of hidden variables Λ. So let $|\Lambda| = \infty$. We have

$$\mathbf{P}^{\text{Mises}}(A) = \lim_{N\to\infty} \nu_N(A) = \lim_{N\to\infty}\lim_{n\to\infty} \sum_{k=1}^{n} \nu_N(\lambda_k) \sum_{j\in j(A)} \nu_N(\lambda'_j/\lambda_k) \,, \tag{2.5}$$

where $\lambda_1, ..., \lambda_n, n = n_N$ are hidden states of particles $s_1, ..., s_N$. On the other hand, we have

$$\mathbf{P}^{\text{com}}(A) == \lim_{n\to\infty} \sum_{k=1}^{n} \lim_{N\to\infty} \nu_N(\lambda_k) \sum_{j\in j(A)} \lim_{N\to\infty} \nu_N(\lambda'_j/\lambda_k) =$$

$$= \lim_{n\to\infty}\lim_{N\to\infty} \sum_{k=1}^{n} \nu_N(\lambda_k) \sum_{j\in j(A)} \nu_N(\lambda'_j/\lambda_k) \,.$$

To obtain the equality $\mathbf{P}^{\text{Mises}}(A) = \mathbf{P}^{\text{com}}(A)$, we have to change the order of limits

$$\lim_{N\to\infty}\lim_{n\to\infty} \;\to\; \lim_{n\to\infty}\lim_{N\to\infty} \,.$$

However, we could not do this in the general case. First of all, as we have already discussed, it may be that $\mathbf{P}^{\text{Mises}}(A) = \lim_{N\to\infty} \nu_N(A)$ exists but some of limits $\lim_{N\to\infty} \nu_N(\lambda_k)$ or $\lim_{N\to\infty} \nu_N(\lambda'_j/\lambda_k)$ do not exists. On the other hand, it may be that, for example, all $\mathbf{p}_k^{\text{Mises}} = \lim_{N\to\infty} \nu_N(\lambda_k) = 0$, i.e.,

$$\lim_{n\to\infty}\lim_{N\to\infty} \sum_{k=1}^{n} \nu_N(\lambda_k) \sum_{j\in j(A)} \nu_N(\lambda'_j/\lambda_k) = 0 \,.$$

But at the same time $\mathbf{P}^{\text{Mises}}(A) \neq 0$. However, it is possible to justify (in some cases) the change of the order of limits with the aid of some regularization procedure.

---

[9]In fact, the $\mathcal{R}$ consists of two regularization procedures: 1) $\mathcal{R}_{\text{fr}}$ gives a regularization ($N \to \infty$) of (in general fluctuating) frequencies; 2) $\mathcal{R}_{\text{conv}}$ gives a regularization ($n \to \infty$) of (in general infinite) average over Λ.

We now consider the ensemble approach to find the origin of negative probabilities. Let us start with following example.

Example 2.2. (Negative distribution of hidden variables) The hidden variable λ has the infinite number of values $\lambda = \lambda_0, ..., \lambda_n, ...$. A statistical ensemble S contains $n(\lambda_l) = 2^l, l = 0, 1, ...,$ particles with $\lambda = \lambda_l$. Let us consider the sub-ensemble $S^{(n)}$ of S which contains all particles with $\lambda \in \{\lambda_0, ..., \lambda_n\}$. Thus $|S^{(n)}| = 1 + \cdots + 2^n = 2^{n+1} - 1$ and $\mathbf{p}_k^{(n)} = \mathbf{P}_{S^{(n)}}(\lambda_k) = \frac{2^k}{2^{n+1}-1}, 0 \leq k \leq n$. The formula of total probability for the ensemble $S^{(n)}$ has the form:

$$\mathbf{P}_{S^{(n)}}(A) = \sum_{k=0}^{n} \mathbf{p}_k^{(n)} \sum_{j \in j(A)} \mathbf{p}_{kj}$$

(here it is assumed that conditional probabilities $\mathbf{p}_{kj}$ depend only on the interaction; they do not depend on n). If $n \to \infty$, then $S^{(n)} \to S$ and $\lim_{n\to\infty} \mathbf{P}_{S^{(n)}}(A) = \mathbf{P}_S(A)$. However, for probabilities $\mathbf{p}_k$ with respect to the ensemble S, we have $\mathbf{p}_k = \lim_{n\to\infty} \mathbf{p}_k^{(n)} = 0$. Thus, in general,

$$\mathbf{P}_S(A) = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbf{p}_k^{(n)} \sum_{j \in j(A)} \mathbf{p}_{kj} \neq \sum_{k=0}^{\infty} \lim_{n\to\infty} \mathbf{p}_k^{(n)} \sum_{j \in j(A)} \mathbf{p}_{kj} = 0\,.$$

We make some formal computations (which, of course, has no meaning in the framework of real analysis). First, we find the 'number of particles' in S :

$$|S| = \sum_{k=0}^{\infty} 2^k = \frac{1}{1-2} = -1\,. \tag{2.6}$$

Then we find probabilities

$$\mathbf{p}_k = \frac{|S(\lambda = \lambda_k)|}{|S|} = -2^k\,. \tag{2.7}$$

Here $\sum_{k=0}^{\infty} \mathbf{p}_k = 1$. Thus we obtained negative ensemble probabilities. We can apply the ensemble formula of total probability to these probabilities:

$$\mathbf{P}_S(A) = \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{j \in j(A)} \mathbf{p}_{kj} \tag{2.8}$$

(at the moment we assume that conditional probabilities are ordinary positive probabilities, $\mathbf{p}_{kj} \geq 0$). Let, for example, $\mathbf{p}_{kj} = q_j \geq 0$ do not depend

on k. Then $\mathbf{P}_S(A) = (\sum_{k=0}^{\infty} \mathbf{p}_k)(\sum_{j\in j(A)} q_j) = \sum_{j\in j(A)} q_j \geq 0$ is ordinary probability (in spite of the presence of negative probabilities). We can also consider k-dependent conditional probabilities $\mathbf{p}_{kj}$. Let, for example, $A = \{\lambda'_0\}$ and $\mathbf{p}_{k0} = 0, k = 2l+1, \ \ \mathbf{p}_{k0} = 1/2^{l+s}, k = 2l$, where $s = 0, 1, ...$ is some (fixed) parameter of the model. Then

$$\mathbf{P}_S(A) = \sum_{l=0}^{\infty} \frac{-2^{2l}}{2^{l+s}} = -\frac{1}{2^s} \sum_{l=0}^{\infty} 2^l = \frac{1}{2^s}$$

is the ordinary probability.

Let $\mathbf{p}_{k0}$ be the same as above and let $\mathbf{p}_{k1} = 1, k = 2l+1, \ \ \mathbf{p}_{k1} = (1 - 1/2^{l+s}), k = 2l$. Set $B = \{\lambda'_1\}$. Then

$$\mathbf{P}_S(B) = \sum_{l=0}^{\infty} \mathbf{p}_{2l+1}\mathbf{p}_{(2l+1)1} + \sum_{l=0}^{\infty} \mathbf{p}_{2l}\mathbf{p}_{(2l)1}$$

$$= -(\sum_{l=0}^{\infty} 2^{2l+1} + \sum_{l=0}^{\infty} 2^{2l}(1 - \frac{1}{2^{l+s}})) = -(-\frac{2}{3} - \frac{1}{3} + \frac{1}{2^s}) = 1 - \frac{1}{2^s} \ .$$

Of course, these 'generalized probabilities' have some properties which are in contradiction with the common probability intuition. For example, let $s = 0$. Then $\mathbf{P}_S(A) = 1$, $\mathbf{P}(B) = 0$ (despite of the fact that $\mathbf{p}_{k1} = 1$ and $\mathbf{p}_k \neq 0, k = 2l+1$).

In Chapter 4 we shall see that all these formal manipulations can be realized on the mathematical level of rigorousness in the p-adic probabilistic framework. In particular, from the p-adic viewpoint probabilities $\mathbf{p}_k = -2^k$ are infinitely small probabilities. Thus in the ensemble S the proportion of systems having the fixed value λ_j of λ is infinitely small. All these infinitely small probabilities must be identified with zero probability in the conventional probability theory.

Example 2.3. (Negative conditional probabilities and negative probabilities for hidden variables). Negative conditional probabilities $\mathbf{p}_{kj}$ may also appear in quite natural statistical ensembles. We assume that the interaction which determines the transition $\lambda_k \to \lambda'_l$ can be represented as a finite chain of steps (trajectory), $(x)_n$ and at each step a particle can have one of two states, 0 or 1. Thus a trajectory of the interaction with n steps has the form $(x)_n = (u_1, ..., u_n), u_j = 0, 1$. In our model we simply assume that the transition $\lambda_k \to \lambda'_l$ is realized via a trajectory of the length l (thus, for fixed l, conditional probabilities $\mathbf{p}_{kl}$ do not depend on k). Consider the statistical

ensemble G_l of trajectories having the length l, where $l = 0, 1, 2...$ (we consider also a 'trajectory' of the length $l = 0$, which describes direct transition $\lambda_k \to \lambda'_0$). Set $G^{(n)} = \cup_{l \leq n} G_l$. Then $|G_l| = 2^l$ and $|G^{(n)}| = 2^{n+1} - 1$ and $|G| = \lim_{n\to\infty} |G^{(n)}| = \sum_{k=0}^{\infty} 2^k = -1$. Thus

$$\mathbf{p}_{kl} = \frac{|G_l|}{|G|} = -2^l .$$

Suppose that as in the above examples $\mathbf{p}_k = -2^k$ and that an experimentally distinguishable state A is determined by values $\lambda'_{2k}, k = 0, 1, ...$, i.e., $A = \{\lambda'_0, ..., \lambda'_{2k}, ...\}$. By the formula of total probability we have

$$\mathbf{P}_S(A) = (\sum_{l=0}^{\infty} -2^l)(\sum_{j=0}^{\infty} -2^{2j}) = \frac{1}{3} .$$

$$\mathbf{P}_S(\bar{A}) = (\sum_{l=0}^{\infty} -2^l)(\sum_{j=0}^{\infty} -2^{2j+1}) = \frac{2}{3} .$$

However, for $A_j = \{\lambda'_j\}$, $\mathbf{P}_S(A_j) = -2^j < 0$.

We shall see in the p-adic framework that such probabilities can be interpreted as infinitely small (but nonzero!) quantities. Thus in this model not only probability to obtain $\lambda = \lambda_j$ for fixed j is infinitely small, but also probability of each transition $\lambda_k \to \lambda'_l$ is infinitely small.

We can easily modify the above example and introduce conditional probabilities $\mathbf{p}_{kj}$ which depend on k.

Example 2.4. (Negative conditional probabilities and positive probabilities for hidden variables) Assume that the interaction which determines the transition $\lambda_k \to \lambda'_l$ can be represented as a chain of the length l of steps (trajectory), $(x)_l$. Assume that at each step a particle can have one of states $d \in D_l = \{d_1, ..., d_l\}$ and each state $d \in D_l$ can appear in a trajectory $(x)_l$ only one time. Thus a trajectory for the transition $\lambda_k \to \lambda'_l$ has the form $(x)_l = (u_1, ..., u_l)$, $u_j \in D_l$, $u_i \neq u_j, i \neq j$, i.e., $(x)_n = \sigma(d_1, ..., d_l)$ is a permutation of elements of the set D_l. It is also assumed that sets of states D_l satisfy the condition of consistency: $D_{l+1} = D_l \cup \{d_{l+1}\}$. We consider now the following statistical ensembles: $G_l, l = 1, 2, ...$ (all trajectories of the length l); $G^{(n)} = \cup_{l=0}^{n} G_l$ (all trajectories of the length $\leq n$); $G = \cup_{l=0}^{\infty} G_l$ (all trajectories of a finite length). Then $|G_l| = l!$, $|G^{(n)}| = \sum_{k=0}^{\infty} k!$. Therefore we obtain that in the framework of real analysis

$$\mathbf{P}_{G^{(n)}}(\lambda'_l/\lambda_k) = \frac{|G_l|}{|G^{(n)}|} \to 0, \quad n \to \infty,$$

i.e., $\mathbf{P}_G(\lambda'_l/\lambda_k) = 0$ in the convectional probability theory. In such a situation (even if hidden variable λ has the ordinary Kolmogorov probability distribution; for example, $\mathbf{p}_k = 1/2^{k+1}, k = 0, 1, ...$) we obtain (of course, only formally) that

$$\mathbf{P}_S(A) = \lim_{n\to\infty} \sum_{k=0}^{n} \mathbf{p}_k \sum_{l\in l(A)} \mathbf{P}_{S_{(n)}}(\lambda'_l/\lambda_k) = \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{l\in l(A)} \lim_{n\to\infty} \mathbf{P}_{G^{(n)}}(\lambda'_l/\lambda_k) = 0.$$

However, if we justify (via some summation procedure) the calculation $|G| = \sum_{k=0}^{\infty} k!$ (in particular, in the p-adic framework), then (nonconventional) probabilities

$$\mathbf{p}_{kl} = \frac{l!}{\sum_{k=0}^{\infty} k!} \neq 0 \tag{2.9}$$

are well defined and the formula of total probability can be applied to these probabilities.

3 Wigner phase-space distribution and negative probability

Even in non-relativistic quantum mechanics negative probabilities creep into the picture. To formulate a conventional (Maxwellian) probability distribution of the coordinates $\mathbf{x}$ and momenta $\mathbf{p}$, similarly to statistical mechanics, is plainly excluded by the corresponding uncertainty relation which prevents at least the simultaneous knowledge of these quantities. Wigner and Szilard, however, found a distribution function which for the first time was applied by Wigner in order to calculate the quantum correction to the gas pressure formula. If a wave function $\psi(x_1, ..., x_n)$, abbreviated by $\psi(\mathbf{x})$, is given, the corresponding Wigner function reads

$$P(\mathbf{x}, \mathbf{p}) = (\pi h)^{-n} \int_{-\infty}^{\infty} d^n\mathbf{y}\bar{\psi}(\mathbf{x}+\mathbf{y})\psi(\mathbf{x}-\mathbf{y}) \exp\{2i(\mathbf{p}, \mathbf{y})/h\}, \tag{3.1}$$

with $\mathbf{x}, \mathbf{y}$ and $\mathbf{p}$ vectors having as many components as has the configuration space of the ψ, namely n; where $(\mathbf{p}, \mathbf{y})$ denotes the scalar product. In order to demonstrate the fundamental features of the Wigner function, relevant for the present purpose, it is sufficient to consider a single particle in linear motion. Thus $n = 1$ and the vector symbols will be dropped henceforth. The

Wigner function exhibits remarkable similarities to a probability distribution in that it leads to the correct probabilities for the coordinates when integrated with respect to the momenta (the integration range is always understood to be $(-\infty, \infty)$ unless indicated otherwise),

$$\int \mathbf{P}(x,p)dp = |\psi(x)|^2, \tag{3.2}$$

and, vice versa, it gives the proper probabilities for the momenta when integrated over the coordinates,

$$\int \mathbf{P}(x,p)dx = (2\pi h)^{-1}|\int dx\psi(x)\exp\{-ipx/h\}|^2 \tag{3.3}$$

Although Wigner calls it the probability function of the simultaneous values for the coordinates and momenta (in more recent papers the notation 'quasi-probability' is adopted) he stresses in the same context, that it cannot really be interpreted in this way "as is clear from the fact, that it may take negative values. But of course this must not hinder the use of it in calculations as an auxiliary function which obeys many relations we would expect from such a probability" [111]. The existence of Wigner functions taking negative values is firmly proved by imposing two very general conditions on $\mathbf{P}$ which can be said to define this type of probability distributions, namely:

(i) $\mathbf{P}(x,p)$ should be a Hermitian form of the state vector $\psi(x)$, i.e., with $\hat{M}(x,p)$ a self-adjoint operator,

$$\mathbf{P}(x,p) = (\psi,\ \hat{M}(x,p)\psi). \tag{3.4}$$

This condition makes $\mathbf{P}(x,p)$ a real number.

(ii) $\mathbf{P}(x,p)$ should give the proper expectation values for all operators which are sums of a function of p and a function of x,

$$\int\int \mathbf{P}(x,p)[f(p)+g(x)]dpdx = \left(\psi,\left[f\left(\frac{h}{i}\frac{\partial}{\partial x}\right)+g(x)\right]\psi\right). \tag{3.5}$$

This condition is a somewhat milder form of (3.2) and (3.3) which properly have to be understood as axioms of the Wigner function and, in any case, must be satisfied. Further, it suffices to consider such ψ which are linear combinations $\psi = a\psi_1 + b\psi_2$ of any two fixed functions, vanishing in certain intervals of x. Now, by requiring

$$\mathbf{P}(x,p) \geq 0 \text{ for all } x \text{ and } p \tag{3.6}$$

for all x and p for every such ψ, Wigner obtains a contradiction which in short runs as follows:

Consider an interval I, inside of which $\psi(x) = 0$ and $g(x) \geq 0$, while $g(x) = 0$ outside and $f(p) \equiv 0$ everywhere. Then (3.5) leads to

$$\int\int \mathbf{P}(x,p)g(x)dpdx = 0. \tag{3.7}$$

Thus

$$\int \mathbf{P}(x,p)g(x)dx = 0 \tag{3.8}$$

for all p (except a set of measure zero).

From (3.6) and the condition imposed on $g(x)$ we obtain (Wigner's lemma): *If $\psi(x)$ vanishes in an interval I, the corresponding $\mathbf{P}(x,p)$ vanishes (except for a set of measure zero) for all values of x in that interval.* Now, consider two functions $\psi_1(x)$ and $\psi_2(x)$ which vanish outside of two non-overlapping intervals I_1, and I_2, respectively. Because of (3.4) $\mathbf{P}(x,p)$ corresponding to $\psi = a\psi_1 + b\psi_2$ will have the form

$$\mathbf{P} = |a|^2\,\mathbf{P}_1 + \bar{a}b\,\mathbf{P}_{12} + a\bar{b}\,\mathbf{P}_{21} + |b|^2\,\mathbf{P}_2. \tag{3.9}$$

By setting $b = 0$ it is obvious that $\mathbf{P}_1$, is the Wigner function of ψ_1, (and $\mathbf{P}_2$ of ψ_2). The meaning of $\mathbf{P}_{12}$ and $\mathbf{P}_{21}$ is less obvious, but we need not bother, because both must be identically zero. This can be seen by considering any interval I' outside I_1. Since, according to the above lemma, $\mathbf{P}$, vanishes almost everywhere in interval I', (3.9) cannot be positive for every choice of a and b unless $\mathbf{P}_{12} = \mathbf{P}_{21} = 0$ outside I_1. The same proof applies to I_2. Thus, instead of (3.9) we have

$$\mathbf{P} = |a|^2\mathbf{P}_1 + |b|^2\mathbf{P}_2 \tag{3.10}$$

almost everywhere. In order to complete the contradiction, let us denote the Fourier transforms of ψ_1 and ψ_2 by $\phi_1(p)$ and $\phi_2(p)$, respectively. Equation (3.3) then reads

$$|a|^2\int \mathbf{P}_1(x,p)dx + |b|^2\int \mathbf{P}_2(x,p)dx$$

$$= |a|^2|\phi_2(p)|^2 + |b|^2|\phi_2(p)|^2 + 2\Re[a\bar{b}\phi_1(p)\phi_2(p)].$$

Since this must be valid for all a and b, we must have identically in p : $\phi_1(p)\bar{\phi}_2(p) = 0$. This is, however, impossible since ϕ_1 and ϕ_2, being Fourier transforms of functions restricted to finite intervals, are analytic functions of their arguments and cannot vanish over any finite interval.

In order to illustrate this result, the Wigner function formalism may be applied to the paradigm of quantum theory, the linear harmonic oscillator (see W. Muckenheim [90]). From its Hamiltonian

$$H(x,p) = p^2/2m + m\omega^2x^2/2 \tag{3.11}$$

and the equation for eigenfunctions of this Hamiltonian:

$$\hat{H}\left(x, \frac{h}{i}\frac{\partial}{\partial x}\right)\psi(x) = E\psi(x). \tag{3.12}$$

It is easy to find the wave function of the ground state

$$\psi_0(x) = (m\omega/h)^{1/4}\exp(-x^2m\omega/2h) \tag{3.13}$$

corresponding to the energy $E_0 = h\omega/2$. Inserting (3.13) in (3.1) and integrating out in y leads to

$$\mathbf{P}_0(x,p) = (\pi h)^{-1}\exp(-x^2m\omega/h - p^2/m\omega h), \tag{3.14}$$

which does not exhibit any anomaly in that it is non-negative and, when integrated with respect to x, supplies the proper distribution of the momentum

$$\int \mathbf{P}_0(x,p)dx = (m\omega\pi h)^{-1/2}\exp(-p^2/m\omega h), \tag{3.15}$$

which is a Gaussian distribution with expectation zero and standard deviation $(\Delta p)^2 = m\omega h/2$. Integrating with respect to p yields, as expected, the square of (3.13),

$$\int \mathbf{P}_0(x,p)dp = (m\omega/\pi h)^{-1/2}\exp(-x^2m\omega/h), \tag{3.16}$$

also a Gaussian distribution with expectation zero and standard deviation $(\Delta x)^2 = h/2m\omega$. Gaussian distributions satisfy Heisenberg's uncertainty relation in its marginal form, i.e., as an equality. From (3.15) and (3.16) we obtain $(\Delta x)(\Delta p)^2 = h/2$. It may also be noted that the distributions of momentum and position are statistically independent, because $\int \mathbf{P}_0 dp \int \mathbf{P}_0 dx = \mathbf{P}_0$.

A fortiori, the covariance coefficient is zero. Clearly, this example does not contradict Wigner's 'negativity proof' because the latter only says that there are state functions for which the corresponding $\mathbf{P}(x,p)$ cannot be everywhere non-negative. One of those is the first excited state of the harmonic oscillator. Using the state function of the first excited level, the same formalism as described above will lead to the corresponding Wigner function.

With H the Hamiltonian of (3.11) and L_n the nth Laguerre polynomial, the Wigner function corresponding to the nth excited state can be expressed by

$$\mathbf{P}_n(x,p) = (\pi h)^{-1}(-1)^n \exp(-2H/h\omega)L_n(4H/h\omega) \tag{3.17}$$

or, using (3.14),

$$\mathbf{P}_n(x,p) = (-1)^n\mathbf{P}_0(x,p)L_n(4H/h\omega). \tag{3.18}$$

As $\mathbf{P}_0$ was found to be non-negative everywhere, we have to examine the remaining expression

$$\mathbf{P}_n/\mathbf{P}_0 = (-1)^n L_n(4H/h\omega). \tag{3.19}$$

The first-order Laguerre polynomial is

$$L_1(u) = 1 - u. \tag{3.20}$$

Hence, $\mathbf{P}_1(x,p)$ goes negative for

$$H \equiv p^2/2m + m\omega^2x^2/2 < h\omega/4. \tag{3.21}$$

Therefore the Wiegner distribution $\mathbf{P}_1(x,p)$ becomes negative only in the extremely small domain (ellips (3.21)). As the energy of the first excited state $E_1 = \frac{3}{2}h\omega$, probability of an energy measurement $\mathbf{P}(E < h\omega/4)$ (where E is the energy of quantum harmonic oscillator) is equal zero. Hence in this example negative values of the Wiegner distribution $\mathbf{P}_1(x,p)$ correspond to events which have zero conventional probability. The use of the Wiegner distribution can be interpreted as a kind of splitting of conventional zero probabilities by using negative numbers (as a class of labels to denote probabilities of events which are identified in the conventional framework with the label '0').

We now consider the Wigner function of the second excited state. The second-order Laguerre polynomial is

$$L_2(u) = 2 - 4u + u^2. \tag{3.22}$$

Using (3.19) we obtain that $\mathbf{P}_2(x,p)$ goes negative for

$$\frac{1}{2}h\omega(1-2^{-1/2}) < H < \frac{1}{2}h\omega(1+2^{-1/2}) , \tag{3.23}$$

As the energy of the second excited state $E_2 = \frac{5}{2}h\omega$, probability of an energy measurement $\mathbf{P}(E < \frac{1}{2}h\omega(1+2^{-1/2}))$ is equal zero. Hence in this example negative values of the Wiegner distribution $\mathbf{P}_1(x,p)$ can be also interpreted as additional labels for probabilities (which are identified with the label '0' in the conventional probability theory) corresponding to events which have zero conventional probability. For $H = 0$ and $H \to \infty$ however, $\mathbf{P}_2(x,p)$ is non-negative.

We will not leave this illustrative example without noting some general features of Wigner functions of the linear harmonic oscillator. From $\lim_{u\to\infty} L_n(u) = (-1)^n u^n$ and (3.17) we find $\mathbf{P}_n$ being positive and asymptotically approaching zero for H going to infinity. In the special case of $H = 0, L_n(0) = n!$ together with (3.17) makes even-order $\mathbf{P}_n$ being positive and odd-order $\mathbf{P}_n$ being negative at $H = 0$.

Most interesting in the present context is, however, that all these Wigner functions of nonzero order unavoidably will take positive as well as negative values. This can easily be seen from the orthogonality relation

$$\frac{1}{n!}\frac{1}{m!}\int_0^\infty e^{-u}L_n(u)L_m(u)du = \delta_{nm}. \tag{3.24}$$

Cohen (see, for example, review [90] for the details) could show that a wide class of probability distribution functions is supplied by the rather general expression

$$\mathbf{P}(x,p) =$$

$$(2\pi)^{-2}\int\int\int f(\theta,\tau)\exp(-i\theta x - i\tau p + i\theta u)\psi^*(u-\tau h/2)\psi(u+\tau h/2)d\theta d\tau du.$$

Herein f is simply a smearing function. By setting $f \equiv 1$, substituting τ by $-2y/h$ and integrating over θ and u, we obtain the original Wigner function (3.1). Other distribution functions may be built with different functions f, if only f satisfies the condition $f(0,\tau) = f(\theta,0) = 1$ in order to yield the correct quantum mechanical marginal distributions.

Cohen imposed the following conditions on a general distribution function $\mathbf{P}(x,p)$: (i) those given by (3.2) and (3.3); (ii) if the quantum mechanical

mean value of the Hermitian operator $\hat{M}$ is $\langle \hat{M} \rangle$, then there should exist a function $g_M(x,p)$ such that

$$\langle \hat{M} \rangle = \int \int g_M(x,p) \mathbf{P}(x,p) dp dx \; ; \tag{3.25}$$

and, for any function K,

$$\langle K(\hat{M}) \rangle = \int \int K(g_M(x,p)) \mathbf{P}(x,p) dp dx \; . \tag{3.26}$$

And he found, that, irrespective of whether $\mathbf{P}$ is positive semidefinite or not, condition (ii) can never be satisfied. The Wigner function $\mathbf{P}$, of the harmonic oscillator, e.g., yields the correct expectation value for the mean energy, but fails to supply the zero-standard deviation, which one should expect from a quantum mechanical energy eigenstate. He concludes: "Of course, it can be argued that the classical formalism does go through as long as we do not insist that the function which must be used to obtain the mean value of a function, K, of g is not identical to $K(g)$. But this would carry us even further from the conceptual basis of classical probability theory than does quantum mechanics itself!".

Finally we note that the equality (3.25) is the direct consequence of the formula of total probability. Let $\hat{M}$ be an orthogonal projector in the Hilbert space of quantum states. It represents the physical observable $M = 0, 1$. Here $\mathbf{P}(M = 1) = \langle \hat{M} \rangle$ and (3.25) is nothing than the formula of total probability for the initial probability distribution $\mathbf{P}(x,p)$ and conditional probabilities $\mathbf{P}(M = 1/(x,p)) = g_M(x,p)$. If we follow to our interpretation of negative probabilities, then we obtain that Hermitian operators represent all physical observables which permit measurements having the property of the statistical stabilization. Non-Hermitian operators represent a new class of physical observables which do not permit measurements with the property of the statistical stabilization. Here we could obtain (for some states) the negative mean value for an observable with positive values.

On the basis of our interpretation of negative probabilities we can finish this section by

Conclusion. *From the frequency viewpoint negative values of Wiegner's probability distribution is nothing than the exhibition of the absence of the statistical stabilization of relative frequencies $\nu_N((x,p) \in U)$ for some domains U of the phase space; from the ensemble viewpoint negative values of Wiegner*

probability distribution is nothing than the exhibition of nonregular structure of infinite statistical ensembles of hidden properties which determine the point (x, p) *of the phase space.*

4 Dirac's world with negative probabilities

The necessity of extended probabilities becomes most distinct if a Lorentz-invariant formulation of quantum theory is attempted. The special role that time plays in non-relativistic theory can, e.g., in the most simple case of particles with no charge and spin, be removed by means of the Klein-Gordon equation which for a single free particle of rest mass m is given by

$$\left(\frac{\partial^2}{\partial x_0^2} + \frac{\partial^2}{\partial x_1^2} + \frac{\partial^2}{\partial x_2^2} + \frac{\partial^2}{\partial x_3^2} + m^2\right)\psi = 0\,,$$

where ($h = c = 1$). Born's notion, however, according to which the square of the wave function has to be interpreted as probability density, necessarily must fail in this context, because $|\psi|^2$ as a scalar violates conservation of total probability. On the other hand, the density proposed by Gordon and Klein

$$\mathbf{P}(x_0, x_1, x_2, x_3) = \frac{1}{2im}\left(\frac{\partial \psi^*}{\partial x_0}\psi - \psi^* \frac{\partial \psi}{\partial x_0}\right), \tag{4.1}$$

satisfies as time component of a four-vector the conservation law, and thus (4.1) is evidently the correct mathematical form to use, but, clearly, it can go negative.

This is not the only difficulty. If the wave function of a plane wave

$$\psi = \exp[-i(p_0 x_0 - p_1 x_1 - p_2 x_2 - p_3 x_3)], \quad p_0 \equiv E,$$

is transformed to the momentum and energy variables, the Gordon-Klein expression (4.1) goes over

$$|\psi(p_0, p_1, p_2, p_3)|^2 p_0^{-1} dp_1 dp_2 dp_3 \tag{4.2}$$

as the probability of the momentum having a value within the small domain $dp_1 dp_2 dp_3$ about a value p_1, p_2, p_3 with the energy having the value p_0, which must be connected with p_1, p_2, p_3 by

$$p_0^2 - p_1^2 - p_2^2 - p_3^2 - m^2 = 0.$$

The weight factor p_0^{-1} appears in (4.2) and makes it Lorentz invariant, since $\psi(p)$ is a scalar - it is defined in terms of $\psi(x)$ to make it so - and the differential element $p_0^{-1}dp_1dp_2dp_3$ is Lorentz invariant. This weight factor may be positive or negative, and makes the probability positive or negative accordingly. Thus the two undesirable things, negative energy and negative probability, always occur together. By our interpretation of negative probabilities one of possible explanations of this fact is that the probability to observe a particle with a negative energy is infinitely small.

Dirac formulates an alternative approach to quantum electrodynamics which allows for a conventional treatment of particles with half-odd integral spin, but unavoidably entails negative probabilities when applied to particles with integral spin, in special cases even demanding probabilities of plus or minus 2, distinctly outside the usual range. On the other hand, this relativistic theory has great advantages over the usual method in that it avoids the *most artificial process of renormalization.* With respect to the latter, Dirac never changed his mind, qualifying it as a 'working rule' and considering its results, in spite of their accuracy, as not reliable. Indeed, the commonly applied method of renormalization is a thing between artificial and nonphysical. We are left between Scylla and Charybdis, in that our equations contain either probabilities as large as plus or minus 2 or electron masses exceeding that of the whole universe. Obviously, also Dirac was very sceptical about those "undesirable things, negative energy and negative probability", but he asserts: "Negative energies and probabilities should not be considered as nonsense. They are well-defined concepts mathematically, like a negative sum of money, since the equations which express the important properties of energies and probabilities can still be used when they are negative. Thus negative energies and probabilities should be considered simply as things which do not appear in experimental results. The physical interpretation of relativistic quantum mechanics that one gets by a natural development of the non-relativistic theory involves these things and is thus in contradiction with experiment. We therefore have to consider ways of modifying or supplementing this interpretation" [34].

To delete the divergences Dirac proposed considering the representation including positive and negative energies. Then to resolve the problem of negative energies he proposed considering operators of emission of photons with negative energy as absorption operators of photons with positive energy. But this picture contains negative probabilities of absorption of any odd number of photons.

Let $A^1(x)$ be operators of the quantum electrodynamics of Heisenberg and Pauli referring to emission and absorption of photons into positive energy

states:

$$A^1(x) = \int\int\int (R_k e^{(k,x)} + \bar{R}_k e^{-(k,x)}) k_0^{-1} dk_1 dk_2 dk_3, \quad (4.3)$$

where $k_0 = +\sqrt{k_1^2 + k_2^2 + k_3^2}$ and R_k being the emission operator and $\bar{R}_k$ the absorption operator. In the same way we introduce the operators $A^2(x)$ referring to the negative energy; there is the representation similar to (4.3) but with $k_0 = -\sqrt{k_1^2 + k_2^2 + k_3^2}$. Dirac considered operators $A^3 = (1/\sqrt{2})(A^1 + A^2)$ which are expended with respect to operators R_k and $\bar{R}_k$ corresponding to positive and negative energies.

The idea was to solve all divergence problems in the symmetric $A^3(x)$ representation. Then we can obtain some information about the $A^1(x)$ representation. But we cannot apply the linear transformation between $A^3(x)$ and $A^1(x)$ representations to the wave function of the $A^3(x)$ representation. There would arise the same divergences. But we can do this with the initial Gibbs ensemble of $A^3(x)$ representation.

It is convenient to consider with $A^3(x)$ additional fields

$$B^3(x) = \frac{1}{\sqrt{2}}(A^1(x) - A^2(x)),$$

which commute with $A^3(x)$, so they are redundant variables. Now let us take B equal to the initial value of A^3. Then for the initial wave function ψ, $(B^3(x) - A^3(x))\psi = 0$ or $\bar{R}_k\psi = 0$ with k_0 either positive or negative. Thus any absorption operator applied to the initial wave function gives the result zero, which means that the corresponding state is one with no photons present.

The following natural interpretation of the wave function at some later time now appears. That part corresponding to m photons of positive energy and n photons of negative energy can be interpreted as corresponding to m photons having been emitted and n photons having been absorbed.

Dirac then considered the momentum representation of $A^3(x)$ and $B^3(x)$ operators. Let k be a momentum–energy vector, $k^2 = 0$, and $\xi_{k\mu}, \xi^*_{k\mu}$ be operators of emission and absorption. There $k_0 = \pm\sqrt{k_1^2 + k_2^2 + k_3^2}$. Then set $\zeta_{k\mu} = \xi_{-k\mu}$ for $k_0 > 0$ and consider the wave function ψ as $\psi = \psi(\xi, \zeta)$, $k_0 > 0$. The following commutation relations take place: $[\xi^*, \xi] = c$ and $[\zeta^*, \zeta] = -c$, $c > 0$.

The variables ξ correspond to the emission of photons of positive energy $k_0 > 0$ and the ζ correspond to the absorption of photons of positive energy

$k_0 > 0$. Let us denote the space of states $\psi(\xi,\zeta)$ by the symbol $\mathcal{H}$. The inner product in $\mathcal{H}$ has the form :

$$(f,g) = \sum_{m,n=0}^{\infty} f_{mn}\bar{g}_{nm} m! c^m n! (-c)^n$$

for the functions

$$f(\xi,\zeta) = \sum_{mn} f_{mn}\xi^m\zeta^n, \quad g(\xi,\zeta) = \sum_{mn} g_{mn}\xi^m\zeta^n.$$

Now for the wave function $\psi(\xi,\zeta)$, normalized by $|\psi|^2 = (\psi,\psi) = 1$, the probability of there having been m photons emitted into momentum and energy state k (corresponding to ξ) and n photons absorbed from this state is

$$\mathbf{P}(m,n) = |\psi_{mn}|^2 c^m m! (-c)^n n!.$$

It gives a negative probability for an odd number of photons having been absorbed. But this statistical interpretation has no meaning in the framework of the ordinary theory of probability. Nevertheless, we can explain the appearance of such 'generalized' probabilities. On one hand, they may appear as a consequence of the violation of the law of large numbers. On the other hand, they demonstrate that Dirac's formalism gives a fine internal structure of theory which could not be described by conventional probabilities.

5 Negative probabilities and localization

One reason for the difficulties with quantum electrodynamics is the general Lorentz condition, according to which the four-divergence of the electromagnetic potential A must vanish

$$\frac{\partial A^0}{\partial x_0} + \frac{\partial A^1}{\partial x_1} + \frac{\partial A^2}{\partial x_2} + \frac{\partial A^3}{\partial x_3} \equiv \partial_\mu A^\mu = 0.$$

A photon density obtained from this continuity equation suffers from the same problems as the Gordon-Klein conserved density (4.1) in that it is not positive semidefinite, or, according to the opinion of the respective referee, it does not exist. This problem might be related to the fact that photons cannot be sharply localized. If they could, we could define the photon density

as the number of photons per unit volume in some arbitrary small volume. However, in a relativistic field, we cannot define such a density.

Therefore it seems that there are two ways for the description of reality: (1) to assume that physical systems could not be localized with arbitrary precision and use Kolmogorov's axiomatic of probability theory; (2) to assume that physical systems could be localized with arbitrary precision, but to change Kolmogorov's axiomatic and create probability theories, where negative probabilities (as well as probabilities which are larger than 1) are mathematically well defined.

If we follow (1), then we have to deny the 'continuous' model of space-time based on real numbers. The system of real numbers **R** describes reality with an infinite precision. Here a physical quantity a is represented by the real number:

$$a = \cdots + \frac{\alpha_{-k}}{m^k} + \cdots + \frac{\alpha_{-1}}{m} + \alpha_0 + \cdots + \alpha_l m^l = \alpha_l \ldots \alpha_0, \alpha_{-1} \ldots \alpha_{-k} \ldots , \quad (5.1)$$

where $\alpha_j = 0, 1, ..., m-1$, and a natural number $m > 1$ gives the scale of a measurement. All digits in (5.1) can be measured (at least theoretically), thus a 'exists with the infinite precision.' It would be natural to consider other number systems based on expansions which are similar to (5.1) and describe reality with a finite precision. Here we could use a system of m-adic numbers $\mathbf{Q}_m$ which is well known in number theory (mainly in the case $m = p$ is a prime number). These are quantities of the form

$$a = \frac{\alpha_{-k}}{m^k} + \cdots + \frac{\alpha_{-1}}{m} + \alpha_0 + \cdots + \alpha_l m^l + \cdots ,$$

where $\alpha_j = 0, 1, ..., m-1$ (thus there is only a finite number of terms corresponding to negative powers of m). A physical formalism based on an m-adic 'finite-precision world' has been developed in [63], [65] (in fact, such a viewpoint is closely connected with the theory of measurements based on nonorthogonal operator valued measures [24], [46], [81]).

If we follow (2), then we can assume that a physical system (in particular, photon) can be localized with an arbitrary precision (i.e., we can still use the real space in quantum theory). However, we could not assume that we should obtain the ordinary (Kolmogorov or Mises) probabilities if we measure statistical distributions corresponding to a 'real localization'.

Our consideration of precision of measurements of physical quantities and negative probabilities can be illustrated by the formalism of quantum theoretical description of radiation. It is given by extending (see the review [90])

the work of Weisskopf and Wigner who calculated the natural linewidth of radiative decay of an excited atom. The corresponding transition amplitude may be rewritten

$$A(E,t) \approx \frac{e^{-t/2} - e^{iEt}}{i/2 - E}$$

with E denoting the difference between actual photon energy and mean state energy E_0 in units of the natural width of the excited state, and t denoting the time interval between excitation and decay in units of the mean lifetime of the state.

It is now very interesting to consider the spectral distribution of photons emitted in finite time intervals. For the time interval $(0,t)$ we have

$$|A(E,t)|^2 = \frac{1}{2\pi}\frac{1 - 2e^{-t/2}\cos(Et) + e^{-t}}{E^2 + 1/4}$$

which undoubtedly is non-negative for every E and t. The spectral distribution emitted at time t, however, $I(E,t) = d|A(E,t)|^2/dt$, entails negative values, as easily can be seen from

$$I(E,t) = \frac{1}{2\pi}\frac{(2E\sin(Et) + \cos(Et))e^{-t/2} - e^{-t}}{E^2 + 1/4}$$

Further, if the quantity $|A(E, t \to \infty)|^2$

$$I_{\mathrm{norm}}(E,t) = \frac{1}{2\pi}\frac{e^{-t}}{E^2 + 1/4}$$

is used to normalize $I(E,t)$, we obtain the normalized decay probability density $\rho(E,t) = I(E,t)/I_{\mathrm{norm}}(E,t)$, which can take on negative values as well as values exceeding unity, and, if integrated over suited domains $\Delta E \Delta t$, small compared to unity $(= h)$, the normalized probability $\rho(E,t)\Delta E \Delta t$, which is an observable quantity, may violate both the lower and the upper limit of Kolmogorov's axiom. These results have been verified by experiments.

As it has been pointed out, if the quantities E and t are measured with extremely high precision, $\Delta E \Delta t < h/2$, then it quite natural that there appear negative probabilities.

Chapter 4

p-adic probability theory

The development of a non-Archimedean (especially, p-adic) mathematical physics [108], [107], [41], [55]–[63], [67]–[69], [4] induced some new mathematical structures over non-Archimedean fields. In particular, probability theory with p-adic valued probabilities was developed in [56], [60], [64], [65], [72], [73]. This probability theory appeared in connection with a model of quantum mechanics with p-adic valued wave functions [57]. The main task of this probability formalism was to present the probability interpretation for p-adic valued wave functions.

The first theory with p-adic probabilities was the frequency theory in which probabilities were defined as limits of relative frequencies $\nu_N = n/N$ in the p-adic topology[1].This frequency probability theory was a natural extension of the frequency probability theory of R. von Mises [86]–[88]. One of the most interesting features of the p-adic frequency theory of probability is the possibility to obtain negative (rational) probabilities as limits of relative frequencies. Thus negative probabilities which has been considered in Chapter 3 can be obtained on the mathematical level of rigorousness as p-adic probabilities. Typically p-adic frequency negative probabilities (as well as probabilities which are larger than 1) appear in the cases of violation of the ordinary Mises statistical stabilization (with respect to the real metric). In fact, in this Chapter we shall only consider a p-adic generalization of Mises' principle of the statistical stabilization. Thus we shall only study a p-adic

[1]The following trivial fact is the cornerstone of this theory: the relative frequencies belong to the field of rational numbers $\mathbf{Q}$; we can study their behaviour not only in the real topology on $\mathbf{Q}$, but also in some other topologies on $\mathbf{Q}$ and, in particular, in the p-adic topologies on $\mathbf{Q}$.

generalization of the notion of the S-sequence. The next natural step is to find a p-adic generalization of Mises' principle of randomness. This problem will be studied in Chapter 6 (on the basis of a p-adic generalization of Martin-Löf's theory of statistical tests).

The next step was the creation of p-adic probability formalism on the basis of a theory of p-adic valued probability measures. It was natural to do this by following the fundamental work of A.N. Kolmogorov [74] in which he had proposed the measure-theoretical axiomatics of probability theory. Kolmogorov used properties of the frequency (Mises) probability (non-negativity, normalization by 1 and additivity) as the basis of his axiomatics. Then he added the technical condition of σ-additivity for using Lebesgue's integration theory. In works [56], [60] we tried to follow A.N. Kolmogorov. p-adic frequency probability has also the properties of additivity, it is normalized by 1 and the set of possible values of this probability is the whole field of p-adic numbers $\mathbf{Q}_p$. Thus it was natural to define p-adic probability as a $\mathbf{Q}_p$-valued measure normalized by 1.

However, it was rather complicated problem to propose a p-adic analogue of the condition of σ-additivity. It is the well known fact that all σ-additive $\mathbf{Q}_p$-valued measures defined on σ-rings are discrete measures [97], [104]. Therefore the creators of non-Archimedean integration theory (A. Monna and T. Springer [89]) did not try to develop abstract measure theory, but they proposed an integration formalism via Bourbaki based on integrals of continuous functions. This integration theory has been used for creating p-adic probability theory in the measure-theoretical framework [60]. The main disadvantage of this probability model is the strong connection with the topological structure of a sample space. This is quite similar to the old probability formalisms of Kolmogorov [75], Frechet [40] and Cramer [23] in which the topological structure of the sample space played the important role.

An abstract theory of non-Archimedean measures has been developed by A. van Rooji [104]. The basic idea of this approach is to study measures defined on *rings* which in principle cannot be extended to measures on σ-rings. This gives the possibility for constructing non-discrete p-adic valued measures. On the other hand, the condition of continuity for measures in [104] implies the σ-additivity in all natural cases[2].

[2]Thus the σ-additivity is not a problem. The problem is find the right domain of definition of p-adic probabilistic measures.

In this Chapter we develop a p-adic probability formalism based on measure theory of [104]. By probabilistic reasons we use the special case of this measure theory: measures defined on *algebras* (such measures have some special properties). However, probabilistic applications stimulate also the development of the general theory of non-Archimedean measures defined on rings. We prove the formula of the change of variables for these measures and use this formula for developing the formalism of conditional expectations for p-adic valued random variables (see also [73]).

The use of p-adic valued probabilistic measures gives the possibility to work on the mathematical level of rigorousness with all signed 'probabilities' (for example, with Wiegner's distribution).

As the fields of p-adic numbers are non-Archimedean there exist infinitely large p-adic numbers (in particular, infinitely large natural numbers) in $\mathbf{Q}_p$. Thus p-adic analysis gives the possibility to use actual infinities and consider statistical ensembles with an infinite number of elements. Probabilities with respect to such ensembles are defined via the standard proportion (used in Chapter 1 for finite ensembles). One of the main features of such ensemble probabilities is the appearance of negative (rational) probabilities (as well as probabilities which are larger than 1). In this approach the origin of such 'pathological' (from the real viewpoint) probabilities is very clear. In particular, we shall see that a large set of negative probabilities is naturally interpreted as a set of infinitely small probabilities (giving the split of the conventional probability 0). We shall also see that a large set of probabilities which are larger than 1 is naturally interpreted as a set of probabilities which are negligibly differ from 1. Other interesting property of p-adic ensemble probability is that the corresponding probabilistic measure is not well defined on a set algebra. The system of events is only a set semi-algebra.

1 Non-Archimedean number systems; p-adic numbers

Here we present a brief introduction to non-Archimedean and, in particular, p-adic analysis (see, for example, [97], [104], [107], [60], [65]).

Let F be a ring[3] (a set where addition, subtraction and multiplication are well defined). Recall that a *norm* is a mapping $|\cdot|_F : F \to \mathbf{R}_+$ satisfying

[3] By a ring we always mean a commutative ring with identity 1.

the following conditions:

$$|x|_F = 0 \Longleftrightarrow x = 0 \text{ and } |1|_F = 1, \tag{1.1}$$

$$|xy|_F \leq |x|_F|y|_F, \tag{1.2}$$

$$|x+y|_F \leq |x|_F + |y|_F. \tag{1.3}$$

The ring F with the norm $|\cdot|_F$ is called a *normed ring.* Set $|F| = \{r \in \mathbf{R}_+ : r = |x|_F, x \in F\}$.

The inequality (1.3) is the well known triangle axiom. A norm is said to be non-Archimedean if the *strong triangle axiom* is valid, i.e.,

$$|x+y|_F \leq \max(|x|_F, |y|_F). \tag{1.4}$$

A ring F with a non-Archimedean norm is said to be a non-Archimedean ring. We shall use the following property of a non-Archimedean norm:

$$|x+y|_F = \max(|x|_F, |y|_F), \text{ if } |x|_F \neq |y|_F. \tag{1.5}$$

In order to prove (1.5) we may assume $|x|_F < |y|_F$. By (1.4) we find $|y|_F \leq \max(|x+y|_F, |x|_F) \leq \max(|x|_F, |y|_F)$. The assumption $|x|_F < |y|_F$ gives $\max(|x|_F, |y|_F) = |y|_F$. Hence $|y|_F = \max(|x+y|_F, |x|_F)$. From $|x|_F < |y|_F$, we deduce $|y|_F = |x+y|_F$. This gives (1.5).

If a norm $|\cdot|_F$ has the property: $|xy|_F = |x|_F|y|_F$, then it is called a *valuation* (sometimes a norm is called a *pseudo-valuation*). A ring F with the valuation $|\cdot|_F$ is called a *valued ring.* The absolute value $|\cdot| \equiv |\cdot|_{\mathbf{R}}$ on the field of real numbers $\mathbf{R}$ is an example of a valuation. This valuation does not satisfy the strong triangle inequality (it satisfies only (1.3)). Valuations and norms with such a property are called Archimedean. Another example of an Archimedean valuation is the absolute value $|\cdot| \equiv |\cdot|_{\mathbf{C}}$ on the field of complex numbers $\mathbf{C}$.

Denote by $\mathbf{Z}(F)$ the ring generated in F by its unity element. If F has zero characteristic (i.e., $n \cdot 1 = 1 + \cdots + 1 \neq 0$ for any $n = 1, 2, ...$), then $\mathbf{Z}(F)$ is isomorphic to the ring of integers $\mathbf{Z}$. Therefore in this case we can consider $\mathbf{Z}$ as a subring of F. In what follows we consider only normed rings F which have zero characteristic.

To illustrate how we can work with the strong triangle inequality we present two simple results.

Proposition 1.1. *Let $|\cdot|_F$ be a non-Archimedean norm. Then $|n|_F \leq 1$ for all elements $n \in \mathbf{Z}$.*

Proof. By the strong triangle inequality (1.4) we have :

$$|n|_F = |1 + \cdots + 1|_F \leq |1|_F = 1.$$

■

Proposition 1.2. *A valuation* $|\cdot|_F$ *is a non-Archimedean valuation if and only if* $|n|_F \leq 1$ *for all elements* $n \in \mathbf{Z}$.

Proof. Let $|n|_F \leq 1$ for all $n = 1, 2, \ldots$ Denote by $\binom{n}{k}$ the binomial coefficients, i.e.,

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}, \quad k \leq n.$$

As these coefficients are integers, $|\binom{n}{k}|_F \leq 1$ for all n and k. Hence we have:

$$|(x+y)^n|_F = |\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}|_F$$

$$\leq \sum_{k=0}^{n} |x|_F^k |y|_F^{n-k} \leq (n+1)(\max |x|_F, |y|_F)^n,$$

i.e.,

$$|x+y|_F \leq \lim_{n\to\infty} (1+n)^{1/n} \max(|x|_F, |y|_F) = \max(|x|_F, |y|_F).$$

■

Let $|\cdot|_F$ be a norm on a ring F. Then the function $\rho_F(x,y) = |x-y|_F$ is a metric on F. It is a translation invariant metric, i.e. $\rho_F(x+h, y+h) = \rho_F(x,y)$. As usual in metric spaces we define 'closed' and 'open' balls in F : $U_r(a) = \{x \in F : \rho_F(x,a) \leq r\}$, $U_r^-(a) = \{x \in F : \rho_F(x,a) < r\}$, $r \in \mathbf{R}_+$. We set $U_r \equiv U_r(0)$. It should be noted that any ball $U_r(a)$, $r \in \mathbf{R}_+$, coincides with some ball $U_s(a)$, $s \in |F|$, $s \leq r$. In what follows we consider only balls $U_r(a)$ with $r \in |F|$. The spheres in F are defined by $S_r(a) = \{x \in F : \rho_F(x,a) = r\}, r \in \mathbf{R}_+$. Of course, if $r \notin |F|$ then $S_r(a) = \emptyset$. Therefore it is meaningful to consider only spheres of radius $r \in |F|$. The normed ring F is *complete* if it is a complete metric space with respect to the metric ρ_F.

Let $|\cdot|_F$ be a non-Archimedean norm. Then the corresponding metric ρ_F satisfies the strong triangle inequality:

$$\rho_F(x,y) \leq \max[\rho_F(x,z), \rho_F(z,y)]. \tag{1.6}$$

Such a kind of metric is called an *ultrametric*. We note that any 'open' or 'closed' ball in an ultrametric space is a simultaneously closed and open

subset. Such sets are called *'clopen'* sets. Spheres in F are also clopen. It seems strange from the point of view of our Eucledean intuition. *The balls U_r are additive subgroups of F :* if $|x|_F, |y|_F \leq r$, then $|x+y|_F \leq \max[|x|_F, |y|_F] \leq r$. Moreover, the ball U_1 is a ring: if $|x|_F, |y|_F \leq 1$ then $|xy|_F \leq |x|_F|y|_F \leq 1$.

We shall continuously use the following simple result.

Lemma 1.1. ('The dream of a bad student') *Let F be a complete non-Archimedean normed ring. The series $\sum_{n=1}^{\infty} a_n$, $a_n \in F$ converges in F if and only if $a_n \to 0$, $n \to \infty$.*

To prove this result we use the Cauchy theorem in complete metric spaces (a sequence $\{S_n\}$ converges iff it is a fundamental sequence, i.e., $|S_n - S_m|_F \to 0$, $n, m \to \infty$) and the estimate $|\sum_{k=n+1}^{m} a_k|_F \leq \max_{n+1 \leq k \leq m} |a_k|_F$.

One of the most important non-Archimedean fields, a system of p-adic numbers $\mathbf{Q}_p$, was constructed by K.Hensel [44]. In fact, it was the first example of a commutative number field (a system where the operations of addition, subtraction, multiplication and division are well defined) which was different from the fields of real and complex numbers. Practically during 100 years p-adic numbers were considered only as objects in pure mathematics. In recent years these numbers have been intensively used in theoretical physics see, for example, the books [107], [60], [65], [52] and papers [108], [41], [3], [53]–[55], [57]–[59], in the theory of probability [56], [65], as well, as in investigations of chaos and dynamical systems [70], [65] and applications to cognitive sciences and psychology [65], [66], [68], [69].

The field of real numbers $\mathbf{R}$ is constructed as the completion of the field of rational numbers $\mathbf{Q}$ with respect to the metric $\rho_R(x,y) = |x-y|$, where $|\cdot|$ is the usual valuation given by the absolute value. The fields of p-adic numbers $\mathbf{Q}_p$ are constructed in a corresponding way, by using other valuations. For any prime number the p-adic valuation $|\cdot|_p$ is defined in the following way. First we define it for natural numbers. Every natural number n can be represented as the product of prime numbers : $n = 2^{r_2}3^{r_3}\cdots p^{r_p}\cdots$. Then we define $|n|_p = p^{-r_p}$, we set in addition $|0|_p = 0$ and $|-n|_p = |n|_p$. We extend the definition of the p-adic valuation $|\cdot|_p$ to all rational numbers by setting for $m \neq 0$: $|n/m|_p = |n|_p/|m|_p$. The completion of Q with respect to the metric $\rho_p(x,y) = |x-y|_p$ is the locally compact field of p-adic numbers $\mathbf{Q}_p$ .It is well known (Ostrovsky's theorem), see [97], that $|\cdot|$ and $|\cdot|_p$ are the only possible valuations on $\mathbf{Q}$. The p-adic valuation satisfies the strong triangle inequality:

$$|x+y|_p \leq \max[|x|_p, |y|_p].$$

Thus the field of p-adic numbers $\mathbf{Q}_p$ is non-Archimedean and the p-adic metric ρ_p is an ultrametric. Thus any p-adic ball $U_r(0)$ is an additive subgroup of $\mathbf{Q}_p$ and the ball $U_1(0)$ is also a ring. It is called the *ring of p-adic integers* and denoted by $\mathbf{Z}_p$.

For any $x \in \mathbf{Q}_p$ we have a unique canonical expansion (converging in the $|\cdot|_p$-norm) of the form

$$x = \alpha_{-n}/p^n + \cdots \alpha_0 + \cdots + \alpha_k p^k + \cdots ,$$

where $\alpha_j = 0, 1, ..., p-1$, are the "digits" of the p-adic expansion. The elements $n \in \mathbf{Z}_p$ have the expansion:

$$n = \alpha_0 + \alpha_1 p + \cdots + a_k p^k + \cdots , \tag{1.7}$$

i.e., they can be identified with sequences of digits

$$n = (\alpha_0, ..., \alpha_k, ...), \alpha_j = 0, 1, ..., p-1. \tag{1.8}$$

If $n \in \mathbf{Z}_p, n \neq 0$, and canonical expansion (1.7) contains only a finite number of nonzero digits α_j, than n is natural number (and vice versa). It is natural to interpret a number $n \in \mathbf{Z}_p$ such that expansion (1.7) contains an infinite number of nonzero digits α_j as an *infinitely large natural number.* Thus the ring of p-adic integers contains actual infinities $n \in \mathbf{Z}_p \setminus \mathbf{N}, n \neq 0$. This is one of the most important features of non-Archimedean number systems (compare with nonstandard numbers [3]) . In section 3 we introduce a partial order structure on $\mathbf{Z}_p$ which extends the standard order structure on $\mathbf{N}$: for $n_1, n_2 \in \mathbf{N}$ $n_1 \leq n_2$ in $\mathbf{N}$ iff $n_1 \leq n_2$ in $\mathbf{Z}_p$. Each finite natural number is less than any infinite number: $n \leq m$ for $n \in \mathbf{N}$ and $m \in \mathbf{Z}_p \setminus \mathbf{N}, m \neq 0$. This order structure will be used to compare p-adic probabilities.

If, instead of a prime number p, we start from an arbitrary natural number $m > 1$, we construct the system of the so called m-adic numbers $\mathbf{Q}_m$ (by completing $\mathbf{Q}$ with respect to the m-adic metric $\rho_m(x, y) = |x - y|_m$). However, this system is not in general a field. There exist in general divisors of zero in $\mathbf{Q}_m$, thus $\mathbf{Q}_m$ is only a ring. Elements of $\mathbf{Z}_m = U_1(0)$ can be identified with sequences (1.8) with the digits $\alpha_k = 0, 1, ..., m-1$. We can also use more complicated number systems corresponding to non-homogeneous scales: $M = (m_1, m_2, ..., m_k, ...)$, where $m_j > 1$ are natural numbers. In this case we obtain the number system $\mathbf{Q}_M$. The elements $x \in \mathbf{Z}_M = U_1(0)$ can be presented as sequences (1.8) with digits $a_j = 0, 1, ..., m_j - 1$. The structure

of $\mathbf{Q}_M$ is rather complicated from the mathematical point of view. In general the number system $\mathbf{Q}_M$ is not a ring. However, $\mathbf{Z}_M$ is always a ring.

Number systems $\mathbf{Q}_m$ and $\mathbf{Q}_M$ can be also used to develop new non-Kolmogorovean probabilistic models. However, the absence of the well developed mathematical formalism does not give such a possibility.

Let K be a non-Archimedean field with the valuation $|\cdot|_K$. Here the function $n \to 1/|n!|_K$ increases (as $|n|_K \leq 1$). The following estimate holds in the field $\mathbf{Q}_p$:

$$(1/np)p^{n/(p-1)} \leq \frac{1}{|n!|_p} \leq p^{(n-1)/(p-1)}. \tag{1.9}$$

This estimate is a consequence of the following mathematical fact:

Lemma 1.2. *Let the natural number n be written in the base p*

$$n = a_0 + a_1 p + ... + a_m p^m, \quad a_j = 0, 1, ..., p-1.$$

Define the sum of the digits of n by $S_n = \sum_{j=0}^m a_j$. Then

$$|n!|_p = p^{(S_n - n)/(p-1)}. \tag{1.10}$$

Proof. There are $[n/p]$ numbers in $\{1, 2, ..., n\}$ that are divisible by p. Here, as usual, $[a]$ is the integer part of a. Then there are $[n/p^2]$ numbers that are divisible by p^2, etc. . By definition $|n!|_p = p^{-\gamma(n)}$, where $\gamma(n) = \sum_{j=0}^m [n/p^j]$. For $j \in 1, 2, ..., m$ we have

$$[n/p^j] = a_j + a_{j+1} p + ... + a_m p^{m-j} = p^{-j} \sum_{i=j}^m a_i p^i.$$

Thus,

$$\gamma(n) = \sum_{j=1}^m p^{-j} \sum_{i=j}^m a_i p^i = \sum_{i=1}^m a_i p^i \sum_{j=1}^i p^{-j}$$

$$= \sum_{i=1}^m a_i p^i \frac{(p^i - 1)}{p^i (p-1)} = (p-1)^{-1} \sum_{i=1}^m a_i (p^i - 1)$$

$$= (p-1)^{-1}(n - S_n).$$

■

By Ostrovsky's theorem the restriction of the valuation $|\cdot|_K$ to $\mathbf{Q}$ is equivalent to one of p-adic valuations: there exists p such that $|x|_K = |x|_p^l$, $l > 0$, for $x \in \mathbf{Q}$. Thus (1.9) implies that

$$a^n \leq \frac{1}{|n!|_K} \leq b^n, \tag{1.11}$$

where $a = a(p, l) > 0$ and $b = b(p, l) > 0$.

The exponent in K is defined by the standard power series $e^x = \sum_{n=0}^{\infty} x^n/n!$. This series converges if $|x|_K < b^{-1}$. In particular, in the p-adic case it converges if $|x|_p < p^{1/(1-p)}$. This is equivalent to $|x|_p \leq r_p$, where $r_p = 1/p$ for $p \neq 2$ and $r_2 = 1/4$. Trigonometric functions over the field K are defined by the standard power series: $\sin x = \sum (-1)^n x^{2n+1}/(2n+1)!$ and $\cos x = \sum x^{2n}/(2n)!$. These series have the same radius of convergence as the series for the exponential function.

2 Frequency Probability Theory

Let us provide a generalization of the von Mises frequency theory of probability. Our main idea is very clear and it is based on the following two remarks: 1) relative frequencies $\nu_N = n/N$ always belong to the field of rational numbers $\mathbf{Q}$; 2) there exist many topologies τ on $\mathbf{Q}$ which are different from the usual real topology τ_R (corresponding to the real metric $\rho_R(x, y) = |x - y|$).

As in ordinary Mises' theory, we also consider infinite sequences

$$x = (x_1, ..., x_N, ...), \quad x_j \in L, \tag{2.1}$$

of observations (here $L = \{\alpha_1, ..., \alpha_k\}$ is a label set). But a new topological principle of the statistical stabilization of relative frequencies is proposed:

the statistical stabilization of relative frequencies $\nu_N(\alpha_i; x)$ can be considered not only in the real topology on the field of rational numbers Q but also in any other topology τ on Q.

This topology is said to be the *topology of statistical stabilization.* Limiting values $\mathbf{P}(\alpha_i) \equiv \mathbf{P}_x^{\tau}(\alpha_i)$ of $\nu_N(\alpha_i; x), i = 1, ..., k$, are said to be τ-probabilities. These probabilities belong to the completion $\mathbf{Q}_\tau$ of $\mathbf{Q}$ with respect to the topology τ. The choice of the topology τ of statistical stabilization is connected with the concrete probabilistic model. Sequence (2.1), for which the principle of statistical stabilization of relative frequencies for the topology τ is valid, is said to be a (S, τ)-sequence (in particular, (S, τ_R)-sequences, where τ_R is the real topology, are ordinary (von Mises) S-sequences which were considered in Chapter 1). At the moment we do not use any τ-analogue of the principle of randomness.

We are mainly interested in the following situation. The real topology τ_R is not a topology of statistical stabilization for the sequence (2.1), but another topology τ is. In this case we cannot consider (2.1) as a von Mises

S-sequence. But there is a new possibility for studying (2.1) as a (S,τ)-sequence.

Set $U_Q = \{q \in \mathbf{Q} : 0 \leq q \leq 1\}$. We denote the closure of the set U_Q in the completion $\mathbf{Q}_\tau$ by U_{Q_τ}. The following theorem is an evident consequence of the topological principle of the statistical stabilization:

Theorem 2.1. *The probabilities* $\mathbf{P}(\alpha_i)$ *belong to the set* U_{Q_τ} *for an arbitrary* (S,τ)*-sequence* x.

As usual, let us consider the algebra F_L of all subsets of L. As in the frequency theory of von Mises we define probabilities $\mathbf{P}(A) = \sum_{\alpha_i \in A} \mathbf{P}(\alpha_i)$ for $A \in F_L$. By Theorem 2.1 the probability $\mathbf{P}(A)$ belongs to the set U_{Q_τ} for every $A \in F_L$.

Theorem 2.2. *Let the completion* $\mathbf{Q}_\tau$ *of* $\mathbf{Q}$ *with respect to the topology of statistical stabilization* τ *be an additive topological group. Then for every* (S,τ)*-sequence* x *the probability is an additive function on* F_L: $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$, $A, B \in F_L$, $A \cap B = \emptyset$.

Here we have used only $\lim(u_N + v_N) = \lim u_N + \lim v_N$ in an additive topological group.

Theorem 2.3. *The probability* $\mathbf{P}(L) = 1$ *for every topology of the statistical stabilization* τ *on* $\mathbf{Q}$.

As in Chapter 1 we define a conditional frequency probability $\mathbf{P}(A/B)$.

Theorem 2.4. *Let* $\mathbf{Q}_\tau$ *be a multiplicative topological group. Then for arbitrary* $A, B \in F_L$, $\mathbf{P}(B) \neq 0$, *the Bayes formula* $\mathbf{P}(A/B) = \mathbf{P}(A \cap B)/\mathbf{P}(B)$ *holds.*

Here we have used $\lim u_N/v_N = \lim u_N/\lim v_N$ if $\lim v_N \neq 0$ in a multiplicative topological group.

However, we may choose the topology of statistical stabilization τ such that $\mathbf{Q}_\tau$ is not an additive group. In this case we obtain non-additive probabilities. Further, $\mathbf{Q}_\tau$ may be not a topological multiplicative group. In this case we have violations of Bayes' formula for conditional probabilities[4]. Moreover, there are possibilities of different combinations of these properties. For example, there exist additive probabilities without Bayes' formula.

Now (following to Kolmogorov) we can present an axiomatics corresponding to the properties of frequency probabilities. Of course, this axiomatics depends on the topology τ. Thus we have an infinite set of axiomatic theories $A(\tau)$. The simplest case (and the one most similar to the Kolmogorov axiomatics) is that $\mathbf{Q}_\tau$ is a topological field. There, by definition, *a* τ*-probability*

[4]A simple realization of Accardi's idea.

is a U_{Q_τ}-valued measure with the normalization condition $\mathbf{P}(\Omega) = 1$. There should be technical restrictions on $\mathbf{P}$ to provide a fruitful theory of integration (compare with Kolmogorov's condition of σ-additivity).

We obtain a large class of non-Komogorov probabilistic models if we choose a metrizable topology τ such that the corresponding metric has the form $\rho_\tau(x, y) = |x - y|_\tau$, where $|\cdot|_\tau$ is a valuation on $\mathbf{Q}$. According to the Ostrovsky theorem, every valuation on $\mathbf{Q}$ is equivalent to the ordinary real absolute value $|\cdot|_R$ or one of the p-adic valuations $|\cdot|_p$. Therefore we may obtain only two classes of probabilistic models: 1) the ordinary theory of probability (with the topology of the statistical stabilization τ_R) ; 2) one of the p-adic valued probabilistic models (with topologies of the statistical stabilization τ_p).

The most interesting property of p-adic probabilities is that $U_{\mathbf{Q}_p} = \mathbf{Q}_p$, see [60]. To prove this fact we need only to show that every $x \in \mathbf{Q}_p$ can be realized as a limit of frequencies $\nu_N = n/N$, where n, N are natural numbers, $n \leq N$. Thus any p-adic number x may be a p-adic probability.

For example, every rational number may be taken as a p-adic probability. There are such 'pathological' probabilities (from the point of view of the usual theory of probability) as $\mathbf{P}(A) = 2$, $\mathbf{P}(A) = 100$, $\mathbf{P}(A) = 5/3$, $\mathbf{P}(A) = -1$. If $p = 1, \mathrm{mod}\ 4$, then $i = \sqrt{-1}$ belongs to $\mathbf{Q}_p$. Thus 'complex quantities' can be obtained as frequency probabilities; for example, $\mathbf{P}(A) = i = \sqrt{-1}$ or $\mathbf{P}(A) = 1 \pm i$.

Thus **negative (and even complex) probabilities can be realized as p-adic frequency probabilities.**

We have presented [60] a large number of statistical models where frequencies oscillate with respect to the real metric ρ_R and stabilize with respect to one of p-adic metrics ρ_p. There p is a parameter of the statistical model. The corresponding statistical simulation was carried out on a computer.

Thus Mises' principle of the statistical stabilization of frequencies can be essentially extended by considering (S, τ)-sequences for topologies τ on $\mathbf{Q}$. It would be natural to extend second Mises' principle, namely, the principle of randomness and introduce an analogue of Mises' collective, namely, a τ-collective. However, I could not obtain any meaningful extension of the principle of randomness for p-adic topologies τ_p. It is still not clear how we can define a class of place selections which would not disturb the p-adic statistical stabilization. On the other hand, it is well known that in ordinary (real) probability theory it is possible to develop the mathematical theory of randomness by using Martin-Löf statistical recursive tests [83]–[85]. In

Chapter 6 we shall follow to P. Martin-Löf and develop a p-adic theory of recursive statistical tests[5].

3 Ensemble Probability

Our interpretation of p-adic numbers

$$N = l_0 + l_1 p + \cdots + l_s p^s + \cdots , \tag{3.1}$$

where $l_s = 0, 1, ..., p-1$, with an infinite number of nonzero digits n_s as infinitely large numbers gives the possibility of considering numerous actual infinities. Therefore we can study ensemble probabilities on ensembles of an infinite volume or consider classical probabilities for an infinite number of equally possible cases.

1. Ensembles of infinite volumes. We shall study some ensembles $S = S_N$ which have a p-dic 'volume' N, where N is the p-adic integer (3.1). If N is finite then S is the ordinary finite ensemble, if N is infinite then S has essentially p-adic structure. Consider a sequence of ensembles M_j having volumes $l_j p^j$, $j = 0, 1, ...$ Set

$$S = \cup_{j=0}^{\infty} M_j.$$

Then $|S| = N$. This split of S will play the crucial role in our probabilistic considerations. Thus S is not just an arbitrary ensemble of the cardinality N. It is an ensemble of the cardinality N constructed via the hierarchical structure corresponding to this split. We may imagine an ensemble S as being the population of a tower $T = T_S$, which has an infinite number of floors with the following distribution of population through floors: population of jth floor is M_j. Set $T_k = \cup_{j=0}^{k} M_j$. This is population of the first $k+1$ floors.

Let $A \subset S$ and let there exist:

$$n(A) = \lim_{k\to\infty} n_k(A), \text{ where } n_k(A) = |A \cap T_k|. \tag{3.2}$$

The quantity $n(A)$ is said to be a *p-adic volume* of the set A.

We define the probability of A by the standard proportional relation:

$$\mathbf{P}(A) \equiv \mathbf{P}_S(A) = \frac{n(A)}{N}. \tag{3.3}$$

[5]Of course, we understood that Martin-Löf's theory does not give the fruitful notion of randomness for an individual sequence of trials.

Denote the family of all $A \subset S$, for which (3.3) exists, by $\mathcal{G}_S$. The sets $A \in \mathcal{G}_S$ are said to be events. Later we shall study some properties of the family of events. First we consider the set algebra F which consists of all finite subsets and their complements.

Proposition 3.1. $F \subset \mathcal{G}_S$.

Proof. Let A be a finite set. Then $n(A) = |A|$ and (3.3) has the form:

$$\mathbf{P}(A) = \frac{|A|}{|S|}. \tag{3.4}$$

Now let $B = \bar{A}$. Then $|B \cap T_k| = |T_k| - |A \cap T_k|$. Hence there exists $\lim_{k\to\infty} |B \cap T_k| = N - |A|$. This equality implies the standard formula:

$$\mathbf{P}(\bar{A}) = 1 - \mathbf{P}(A). \tag{3.5}$$

■

In particular, we have : $\mathbf{P}(S) = 1$.

Proposition 3.2. *Let $A_1, A_2 \in \mathcal{G}_S$ and $A_1 \cap A_2 = \emptyset$. Then $A_1 \cup A_2 \in \mathcal{G}_S$ and*

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2). \tag{3.6}$$

Proposition 3.3. *Let $A_1, A_2 \in \mathcal{G}_S$. The following conditions are equivalent:*

$$1) A_1 \cup A_2 \in \mathcal{G}_S; \quad 2) A_1 \cap A_2 \in \mathcal{G}_S;$$

$$3) A_1 \setminus A_2 \in \mathcal{G}_S; \quad 4) A_2 \setminus A_1 \in \mathcal{G}_S.$$

There are standard formulas:

$$\mathbf{P}(A_1 \cup A_2) = \mathbf{P}(A_1) + \mathbf{P}(A_2) - \mathbf{P}(A_1 \cap A_2); \tag{3.7}$$

$$\mathbf{P}(A_1 \setminus A_2) = \mathbf{P}(A_1) - \mathbf{P}(A_1 \cap A_2). \tag{3.8}$$

Proof. We have $n_k(A_1 \cup A_2) = n_k(A_1) + n_k(A_2) - n_k(A_1 \cap A_2)$. Therefore, if, for example, $A_1 \cap A_2 \in \mathcal{G}_S$ then there exists a limit of the right hand side. It implies $A_1 \cup A_2 \in \mathcal{G}_S$ and (3.7) holds. Other implications are proved in the same way. ■

Corollary 3.1 *The family $\mathcal{G}_S$ is a semi-algebra.*

In general $A_1, A_2 \in \mathcal{G}_S$ does not imply $A_1 \cup A_2 \in \mathcal{G}_S$. To show this, by Proposition 3.3 it suffices to find $A_1, A_2 \in \mathcal{G}_S$ such that $A_1 \cap A_2 \notin \mathcal{G}_S$ It is easy to do: let $A_1, A_2 \in \mathcal{G}_S$ are such that $|A_1 \cap A_2 \cap M_l| = 1$ for nonempty

M_l (there is only one element $x \in A_1 \cap A_2$ on each nonempty floor). If N is infinite then $\lim_{k\to\infty} n_k(A_1 \cap A_2)$ does not exist. Thus

$\mathcal{G}_S$ is not a set algebra.

It is closed only with respect to a finite unions of sets which have empty intersections. However, $\mathcal{G}_S$ is not closed with respect to countable unions of such sets: in general ($A_j \in \mathcal{G}_S$, $j = 1, 2, ...,$ $A_i \cap A_j = \emptyset$, $i \neq j$,) does not imply $\cup_{j=1}^{\infty} A_j \in \mathcal{G}_S$. The natural additional assumptions (A) $\sum_{j=1}^{\infty} \mathbf{P}(A_j)$ converges in $\mathbf{Q}_p$ or (more strong assumption), (B) $\sum_{j=1}^{\infty} |\mathbf{P}(A_j)|_p < \infty$, also do not imply $A \in \mathcal{G}_S$.

Example 3.1. Let $m = 2$, $N = -1 = 1 + 2 + 2^2 + \cdots + 2^n + \cdots$. Suppose that the sets A_j have the following structure: $|A_j \cap M_{3(j-1)}| = 1$, $|A_j \cap M_{3j-1}| = 2^{3j-1} - 1$ and $A_j \cap M_i = \emptyset$, $i \neq 3(j-1)$, $3j-1$, i.e., the set A_j is located on two floors of the tower T. In particular, $A_i \cap A_j = \emptyset$, $i \neq j$. As $A_j \in F$, then $A_j \in \mathcal{G}_S$; the probability $\mathbf{P}(A_j) = -2^{3j-1}, j = 1, 2\ldots$. The series $\sum_{j=1}^{\infty} |\mathbf{P}(A_j)|_2 < \infty$. We show that $A = \cup_{j=1}^{\infty} A_j \notin \mathcal{G}_S$. We have:

$$n_{3(j-1)}(A) = |A_j \cap T_{3(j-1)}| + | \cup_{s=1}^{j-1} A_s \cap T_{3(j-1)}| = 1 + \gamma,$$

where $|\gamma|_2 < 1$. Thus $|n_{3(j-1)}(A)|_2 = 1$. But $|n_{3j-1}(A)|_2 < 1$.

We note the following useful formula for computing probabilities:

$$\mathbf{P}(A) = \sum_{j=0}^{\infty} \mathbf{P}(A \cap M_j)$$

(probability to find in the tower T an inhabitant $\mathcal{I}$ with the property A is equal to the sum of probabilities to find an inhabitant with this property on the fixed floor).

Definition 3.1. *The system $\mathcal{P} = (S, \mathcal{G}_S, \mathbf{P}_S)$ is called a p-adic ensemble probability space for the ensemble S.*

If N is a finite natural number then we obtain the ensemble probability space which was considered in Chapter 1 (with $\mathcal{G}_S = F_S$). In fact, any ensemble probability space $\mathcal{P}$ can be approximated by ensemble probability spaces $\mathcal{P}_k$ having ensembles of finite volumes. Set

$$n_k = l_0 + l_1 p + \cdots + l_k p^k$$

for N which has the expansion (3.1). Let l_s be the first nonzero digit in (3.1). Consider finite ensembles S_{n_k}, $|S_{n_k}| = n_k$ ($k = s, s+1, ...$), and ensemble probability spaces $\mathcal{P}_{n_k} = (S_{n_k}, \mathcal{G}_{S_{n_k}}, \mathbf{P}_{S_{n_k}})$. There $\mathcal{G}_{S_{n_k}}$ coincides with the

algebra $F_{S_{n_k}}$ of all subsets of the finite ensemble S_{n_k} and definition (3.3) of ensemble probability coincides with the definition of Chapter 1:

$$\mathbf{P}_{S_{n_k}}(A) = \frac{|A|}{|S_{n_k}|}, \quad A \in F_{S_{n_k}}. \tag{3.9}$$

We identify S_{n_k} with the population of the first $k+1$ floors of the tower T_S.

Proposition 3.4. *Let $A \in \mathcal{G}_S$. Then*

$$\mathbf{P}_S(A) = \lim_{k\to\infty} \mathbf{P}_{S_{n_k}}(A \cap S_{n_k}). \tag{3.10}$$

To prove (3.10) we have only used that $\mathbf{Q}_p$ is a topological group. This approximation depends essentially on the rule of a measurement, which is defined by the sequence $\{n_k\}$ which gives an approximation of the infinite ensemble S by finite ensembles $\{S_{n_k}\}$. In principle the change of this rule may change the limiting result (see [60] for the details).

Proposition 3.5. (The image of ensemble probability). *The probability* $\mathbf{P}$ *maps $\mathcal{G}_S$ into the ball $U_{r_S}(0)$, where $r_S = 1/|N|_p$.*

To study conditional probabilities we have to extend the notion of the p-adic ensemble probability to consider more general ensembles.

Let S be the population of the tower T_S with an infinite number of floors M_j, $j = 0, 1, \ldots$, and the following distribution of population: there are m_j elements on the jth floor, $m_j \in \mathbf{N}$ and the series $\sum_{j=1}^{\infty} m_j$ converges in $\mathbf{Z}_p$ to a nonzero number $N = |S|$. We define the p-adic ensemble probability of a set $A \subset S$ by (3.2), (3.3); $\mathcal{G}_S$ is the corresponding family of events. It is easy to check that Propositions 3.1–3.5 hold for this more general ensemble probability.

Let $A \in \mathcal{G}_S$ and $\mathbf{P}(A) \neq 0$. We can consider A as a new ensemble with the p-adic hierarchical structure $A = \cup_{j=0}^{\infty} M_{Aj}$, where $M_{Aj} = A \cap M_j$, and introduce the corresponding family of events $\mathcal{G}_A$.

Proposition 3.6. (Conditional probability). *Let $A \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$ and $B \in \mathcal{G}_A$. Then $B \in \mathcal{G}_S$ and Bayes' formula*

$$\mathbf{P}_A(B) = \frac{\mathbf{P}_S(B)}{\mathbf{P}_S(A)} \tag{3.11}$$

holds true.

Proof. The tower T_A of the A has the following population structure: there are M_{Aj} elements on the jth floor. In particular, $T_{Ak} = T_k \cap A$. Thus

$$n_{Ak}(B) = |B \cap T_{Ak}| = |B \cap T_k| = n_k(B) \tag{3.12}$$

for each $B \subset A$. Hence the existence of $n_A(B) = \lim_{k\to\infty} n_{Ak}(B)$ implies the existence of $n_S(B) = \lim_{k\to\infty} n_k(B)$. Moreover, $n_S(B) = n_A(B)$. Therefore,

$$\mathbf{P}_A(B) = \frac{n_A(B)}{n_S(A)} = \frac{n_A(B)/|S|}{n_S(A)/|S|}.$$

■

By (3.12) we obtain the following consequence:

Corollary 3.2. *Let $A, B \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$, and $B \subset A$. Then $B \in \mathcal{G}_A$.*

Thus we obtain

$$\mathcal{G}_A = \{B \in \mathcal{G}_S : B \subset A\}.$$

Let $A, B, A \cap B \in \mathcal{G}_S, \mathbf{P}(A) \neq 0$. We set by definition $\mathbf{P}_A(B) = \mathbf{P}_A(A \cap B)$. Then

$$\mathbf{P}_A(B) = \frac{\mathbf{P}_S(B \cap A)}{\mathbf{P}_S(A)}. \tag{3.13}$$

If we set $\mathbf{P}_A(B) = \mathbf{P}(B/A)$ and omit the index S for the probabilities for an ensemble S, then we obtain Bayes' formula.

Remark 3.1. We have discussed many times the domain of applications of Bayes' formula. This question has the exact and simple mathematical answer in the p-adic ensemble probability theory. We can use Bayes' formula for events A and B iff $A \cap B$ is also the event, i.e., $A \cap B \in \mathcal{G}_S$.

Remark 3.2. It is important for our physical considerations that $\mathcal{G}_S$ is not a set algebra and $\mathbf{P}_S$ can in principle take any value $x \in U_{r_S}$. The manipulations which were used to prove Bell's inequality (Chapter 2) are not legal for the ensemble probability space $\mathcal{P} = (S, \mathcal{G}_S, \mathbf{P}_S)$. For instance, if there are tree sets $B_\phi, B_\theta, B_0 \in \mathcal{G}_S$, then in principle it may be that $B_\phi \cap B_\theta$, $B_\phi \cap B_0$, $B_0 \cap B_\theta \in \mathcal{G}_S$, but $B_\phi \cap B_\theta \cap B_0 \notin \mathcal{G}_S$. Moreover, probabilities can in principle be negative. In this case we cannot use the standard estimate for Kolmogorov probabilities.

2. The rules for working with p-adic probabilities. One of the main tools of the ordinary theory of probability is based on the order structure on the field of real numbers $\mathbf{R}$. It gives the possibility of comparing probabilities of different events; events E with probabilities $\mathbf{P}(E) \ll 1$ are considered as negligible and events E with probabilities $\mathbf{P}(E) \approx 1$ are considered as practically certain. However, the use of these relations in concrete applications is essentially based on our (real) probability intuition. What is a large probability? What is a small probability? Moreover, it is not easy to compare two arbitrary probabilities. For instance, do you prefer to win with the

probability $\mathbf{P}(E_1) = \frac{11}{17}$ or $\mathbf{P}(E_2) = \frac{13}{19}$. Formally, because $\mathbf{P}(E_1) < \mathbf{P}(E_2)$ it would be better to choose E_2. But in practice this choice does not give many advantages. Thus ordinary probability intuition is based more on centuries of human experiment than on exact mathematical theory.

If we want to work with p-adic probabilities we have to develop some kind of a p-adic probability intuition. However, there arises a mathematical problem which does not give the possibility of generalizing the real scheme directly. This is the absence of an order structure on $\mathbf{Q}_p$. Of course, we can also do something without an order structure. For example, we can classify (split) different events with the aid of their p-adic probabilities. For instance, it works sufficiently successful in the frequency probability theory. If there are two sequences x and y (generated by some statistical experiment) which are not S-sequences in the ordinary von Mises' frequency theory, then we could not split properties of x and y. Both these sequences seem to be totally chaotic from the real point of view. However, if they are (S, τ_p)-sequences, then it would be possible to classify them with the aid of p-adic probability distributions, $\mathbf{P}_x(\alpha_i), \mathbf{P}_y(\alpha_i)$. In the ensemble approach different p-adic probabilities, $\mathbf{P}_S(E_1) \neq \mathbf{P}_S(E_2)$, mean that the events E_1 and E_2 have different p-adic volumes.

However, we could do much more with p-adic probabilities by using the partial order structure which exists on the ring of p-adic integers.

($\mathcal{O}$) Let $x = x_0x_1...x_n...$ and $y = y_0y_1...y_n...$ be the canonical expansions of two p-adic integers $x, y \in \mathbf{Z}_p$. We set $x < y$ if there exists n such that $x_n < y_n$ and $x_k \leq y_k$ for all $k > n$.

This partial order structure on $\mathbf{Z}_p$ is the natural extension of the standard order structure on the set of natural numbers $\mathbf{N}$. It is easy to see that $x < y$ for any $x \in \mathbf{N}$ and $y \in \mathbf{Z}_p \setminus \mathbf{N}$, i.e., any finite natural number is less that any infinite number. But we could not compare any two infinite numbers.

Example 3.2. Let $p = 2$ and let $x = -1/3 = 10101....1010...$, $z = -2/3 = 0101....0101...$ and $y = -16 = 0001...1111...$. Then $x < y$ and $z < y$, but the numbers x and z are incompatible.

It is important to remark that there exists the maximal number $N_{max} \in \mathbf{Z}_p$. It is easy to see:

$$N_{max} = -1 = (p-1) + (p-1)p + \cdots + (p-1)p^n + \cdots .$$

Therefore the ensemble S_{-1} is the largest ensemble which can be considered in the p-adic framework.

Remark 3.3. It seems to be natural to suppose that the volume of the ensemble increases with the increase of p, i.e., $|S^p_{-1}| < |S^q_{-1}|, p < q$.

Proposition 3.7. *Let $N \in \mathbf{Z}_p, N \neq 0$. Then $S_N \in \mathcal{G}_{S_{-1}}$ and*

$$\mathbf{P}_{S_{-1}}(S_N) = \frac{|S_N|}{|S_{-1}|} = -N. \tag{3.14}$$

Corollary 3.3. *Let $N \in \mathbf{Z}_p, N \neq 0$. Then $\mathcal{G}_{S_N} \subset \mathcal{G}_{S_{-1}}$ and probabilities $\mathbf{P}_{S_N}(A)$ are calculated as conditional probabilities with respect to the sub-ensemble S_N of ensemble S_{-1} :*

$$\mathbf{P}_{S_N}(A) = \mathbf{P}_{S_{-1}}(A/S_N) = \frac{\mathbf{P}_{S_{-1}}(A)}{\mathbf{P}_{S_{-1}}(S_N)},\ A \in \mathcal{G}_{S_N}. \tag{3.15}$$

But $A \in \mathcal{G}_{S_{-1}}$ does not imply $A \cap S_N \in \mathcal{G}_{S_N}$.

By Corollary 3.3 we can, in fact, restrict our considerations to the case of the maximal ensemble S_{-1}. Therefore we shall study this case $S \equiv S_{-1}$.

The (partial) order $\mathcal{O}$ on the set of p-adic integers $\mathbf{Z}_p$ gives the possibility to compare p-adic volumes $n(A)$ of sets $A \in \mathcal{G}_S$. It is natural to say that probability $\mathbf{P}(B)$ is larger than probability $\mathbf{P}(A)$ if the p-adic volume $n(B)$ of B is larger than the p-adic volume $n(A)$ of A. Thus we obtain the following (partial) order on the set of probabilities:

($\tilde{\mathcal{O}}$) $\mathbf{P}(B) > \mathbf{P}(A)$ iff $n(B) > n(A)$.

We use th same symbols $>, <$ for this new order on $\mathbf{Z}_p$. We hope that the reader would not mix these two orders on $\mathbf{Z}_p$: $\mathcal{O}$-order is used to compare p-adic volumes, $\tilde{\mathcal{O}}$-order is used to compare probabilities. For example, let $p = 2$ and let $n(B) = -2(= 011...1...), n(A) = -3(= 1011...1...)$. Then $n(B) > n(A)$ (with respect to $\mathcal{O}$) and consequently $\mathbf{P}(B) = 2 > \mathbf{P}(A) = 3$ (with respect to $\tilde{\mathcal{O}}$)

We study some properties of probabilities.

(1) As we have only a partial order structure we cannot compare probabilities of arbitrary two events A and B.

(2) As $x \leq -1$ with respect to $\mathcal{O}$ for any $x \in \mathbf{Z}_p$, we have $\mathbf{P}(A) \leq 1 = \mathbf{P}(S)$ for any $A \in \mathcal{G}_S$.

(3) As $x \geq 0$ with respect to $\mathcal{O}$ for any $x \in \mathbf{Z}_p$, we have $\mathbf{P}(A) \geq 0$ for any $A \in \mathcal{G}_S$.

To illustrate further properties of p-adic probabilities, we shall use the third order structure, namely, the usual real order structure on the set $\mathbf{Z}_p \cap \mathbf{Q}$. In this case we shall say r-increase or r-decrease. This r-order on $\mathbf{Z}_p \cap \mathbf{Q}$ has

no probabilistic meaning. We consider this order, because we want to use the 'real intuition' to imagine the location of rational probabilities $\mathbf{P}(A), A \in \mathcal{G}_S$, on the real line. We shall use the symbols $[a,b], ..., (a,b)$ for corresponding intervals of the real line. For example, let $p = 2$ and let $\mathbf{P}(B) = 2$ and $\mathbf{P}(A) = 3$. Then $\mathbf{P}(B) > \mathbf{P}(A)$, but from the viewpoint of the r-order $\mathbf{P}(B)$ is less than $\mathbf{P}(A)$.

(4) Set $F^f = \{A \in \mathcal{G}_S : n(A) \in \mathbf{N}\}$. [6]

The restriction of the order $\mathcal{O}$ on the set of natural numbers $\mathbf{N}$ coincides with the standard (real) order on $\mathbf{N}$. Thus $n(A) < n(B),\ A, B \in F^f$, iff the natural number $n(A)$ is less than the natural number $n(B)$. This implies (by definition of the order $\tilde{\mathcal{O}}$ on the set of probabilities) that $\mathbf{P} : F^f \to (-\infty, 0) \cap \mathbf{Z}$ and $\mathbf{P}(A)$ is increasing if $\mathbf{P}(A)$ is r-decreasing. Therefore, for example, probabilities $\mathbf{P}(A) = -1$ or -3 are rather small with respect to probabilities $\mathbf{P}(B) = -100$ or -300.

(5) Set $\bar{F}^f = \{B = \bar{A} : A \in F^f\}$ (in particular, $\bar{F}^f$ contains complements of all finite subsets of Ω). Then $\mathbf{P} : \bar{F}^f \to \mathbf{N}$ and $\mathbf{P}(B)$ is decreasing if $\mathbf{P}(B)$ is r-increasing. Therefore, for example, probabilities $\mathbf{P}(E) = 100$ or 200 are rather small with respect to probabilities $\mathbf{P}(C) = 1$ or 2.

We can use these rules for conditional probabilities. For example, let $\mathbf{P}(B) = 100,\ \mathbf{P}(B') = 200,\ \mathbf{P}(A) = 2$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 50 > \mathbf{P}(B'/A) = 100$.

By (4) and (5) we can work with probabilities belonging to $F^f \cup \bar{F}^f$.

(6) Now consider events $A \notin F^f \cup \bar{F}^f$. We can develop our intuition only by examples.

Example 3.3. Let $p = 2$. Let $|A \cap M_{2k}| = 2^{2k}$ and $A \cap M_{2k+1} = \emptyset,\ k = 0, 1, ...$. Then $n(A) = -1/3(= 1010...10...)$ and $\mathbf{P}(A) = 1/3$. Let $B \subset A$ and $B \cap M_{4k} = A \cap M_{4k},\ B \cap M_j = \emptyset,\ j \neq 4k$. Then $n(B) = -1/15(= 100010001...10001...)$ and $\mathbf{P}(B) = 1/15$. It is evident that $-1/15 < -1/3$ in $\mathbf{Z}_2$. Hence $\mathbf{P}(B) = 1/15 < \mathbf{P}(A) = 1/3$.

Thus it seems to be that the probabilistic order relation on the set $[0,1] \cap \mathbf{Q}$ coincides with the standard real order. Moreover, it seems to be reasonable to use this relation also in the case where the numbers $n(A)$ and $n(B)$ are incompatible in $\mathbf{Z}_2$ [7].

[6] In particular, F^f contains all finite subsets of S. The F^f contains also some infinite subsets $A \in \mathcal{G}_S$ which have finite p-adic volumes. For example, let $|A \cap T_k| = 1 + p^k, k = 1, 2, ...$ ($1 + p^k$ inhabitants of the first $(k+1)$ floors have the property A). Then $n(A) = 1$ and hence $A \in F^f$.

[7] However, probably it is the wrong extrapolation and we must assume existence of

Example 3.4. Let p and A be the same as above. Let $|C \cap M_{2k+1}| = 2^{2k+1}$, $C \cap M_{2k} = \emptyset$, $k = 0, 1, \ldots$. Then $n(C) = -2/3$ and $\mathbf{P}(C) = 2/3$. The numbers $n(A) = -1/3$ and $n(C) = -2/3$ are incompatible in $\mathbf{Z}_2$. But heuristically it seems to be evident that we can use the r-order structure on $[0, 1]$ to compare the probabilities of the events A and C. Therefore the probability of $\omega \in C$ is two times larger than the probability $\omega \in A$. These heuristic reasons were also confirmed by some frequency statistical models, see [60] for the details.

Further we have that a probability $x \in (-\infty, 0) \cap \mathbf{Z}$ is practically negligible with respect to any probability $y \in (0, 1] \cap \mathbf{Q}$. The intuitive argument is the following. A probability $\mathbf{P}(A) \in (-\infty, 0) \cap \mathbf{Z}$ is probability of an event A with a finite p-adic volume in the infinitely large ensemble S. Probability $\mathbf{P}(A) \in (0, 1] \cap \mathbf{Q}$ is probability of an event A with an infinite p-adic volume in the infinitely large ensemble S.

Therefore, p-adics gives the possibility to split probability 0 to a set of probabilities, $0 \to D_0^+$; in particular, $(-\infty, 0) \cap \mathbf{Z} \subset D_0^+$.

Remark 3.4. A probability $\mathbf{P}$ on a Boolean algebra $\mathcal{A}$ is non-degenerated: $\mathbf{P}(A) = 0, A \in \mathcal{A}$ iff $A = \emptyset$. The p-adic split of probability 0 can be considered as a step in the direction to Boolean probabilities. The set of new labels D_0^+ gives the possibility to split many probabilities which must be equal to probability 0 from the viewpoint of real analysis. However, we still have not obtained a Boolean probability. There are numerous events $A \in \mathcal{G}_S, A \neq \emptyset$, which have probability 0. For example, let $|A \cap T_k| = p^k, k = 1, 2, \ldots$. Then $\mathbf{P}(A) = 0$.

We can also use these rules for conditional probabilities. For example, let $\mathbf{P}(B) = 1/15 < \mathbf{P}(B') = 2/15$, $\mathbf{P}(A) = 1/5$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 1/3 < \mathbf{P}(B'/A) = 2/3$. Moreover, for example, let $\mathbf{P}(B) = -1 < \mathbf{P}(B') = -5$, $\mathbf{P}(A) = -100$ and $B, B' \subset A$. Then $\mathbf{P}(B/A) = 1/100 < \mathbf{P}(B'/A) = 1/20$. Thus the r-order structure on $(0, 1] \cap \mathbf{Q}$ reproduces the rule (4).

Proposition 3.8. *If* $\mathbf{P}(B) \in \mathbf{N}$, *then* $n(\bar{B}) \in \{0\} \cup \mathbf{N}$; *if* $\mathbf{P}(B) \in (0, 1) \cap \mathbf{Q}$ *then* $n(\bar{B}) \in \mathbf{Z}_p \setminus \mathbf{N}$.

Proof. If $k = \mathbf{P}(B) \in \mathbf{N}$, then $n(B) = -k, k = 1, 2, \ldots$, and $n(\bar{B}) = -1 + k$. If $a = \mathbf{P}(B) \in (0, 1) \cap \mathbf{Q}$ then $n(B) = -a$ and $n(\bar{B}) = a - 1 \notin \mathbf{N}$. ∎

Thus if $\mathbf{P}(B) \in \mathbf{N}$, then the set $\bar{B}$ has a finite p-adic volume, $n(\bar{B})$. On the other hand, if $\mathbf{P}(B) \in (0, 1) \cap \mathbf{Q}$, then the set $\bar{B}$ has an infinite p-adic volume, $n(\bar{B})$. It is natural to assume that probability $\mathbf{P}(B) \in \mathbf{N}$ is larger

events with incompatible probabilities.

than any probability $\mathbf{P}(C) \in (0,1) \cap \mathbf{Q}$.

Therefore, p-adics gives the possibility to split probability 1 to a set of probabilities, $1 \to D_1^-$. In particular, $\mathbf{N} \subset D_1^-$. However, the probability 1 is still not totally split. There are numerous events $A \neq \emptyset$ with $\mathbf{P}(A) = 1$. For example, let $|A \cap M_k| = p^{[(k+1)/2]} - 1, k = 1, 2, \ldots$ (here $[x]$ denotes the integer part of x). Then $n(A) = -1$ and $\mathbf{P}(A) = 1$. But $\bar{A} \neq \emptyset$.

We can also split all probabilities $x = \mathbf{P}(A) \in (0,1) \cap \mathbf{Q}$.

Let $A \in \mathcal{G}_S, x = \mathbf{P}(A) \in (0,1) \cap \mathbf{Q}$, $C \in F^f, A \cap C = \emptyset$, and let $B = A \cup C$. Then $\lambda = \mathbf{P}(B) = \mathbf{P}(A) + \mathbf{P}(C) = x - k$, where $\mathbf{P}(C) = -k, k \in \mathbf{N}$. As the p-adic volume of the set C is finite (and the ensemble S is infinite) probability $\mathbf{P}(C) = -k$ is infinitely small. Thus the probability x can be split in a set of probabilities D_x^+. Each probability $\lambda \in D_x^+$ is larger than probability x and probability $\Delta = \lambda - x = -k$ is infinitely small.

Let $B \in \mathcal{G}_S, C \in F^f, B \cap C = \emptyset$, and let $A = B \cup C, x = \mathbf{P}(A) \in (0,1) \cap \mathbf{Q}$. Then $\lambda = \mathbf{P}(B) = \mathbf{P}(A) - \mathbf{P}(C) = x + k$, where $\mathbf{P}(C) = -k, k \in \mathbf{N}$, is infinitely small probability. Thus the probability x can be split in a set of probabilities D_x^-. Each probability $\lambda \in D_x^-$ is less than probability x and probability $\Delta = x - \lambda = -k$ is infinitely small.

Thus probability x is split in a set of probabilities $D_x = D_x^- \cup D_x^+$.

We now consider probabilities with respect to an ensemble S_N for an arbitrary $N \in \mathbf{Z}_p, N \neq 0$. By using formula (3.15) we can translate to the general case results obtained for the ensemble $S = S_{-1}$. In the general case probability 0 is split in a set D_0^+ which contains the set $\{\lambda = \frac{k}{N} : k \in \mathbf{N}\}$; probability 1 is split in a set D_1^- which contains the set $\{\lambda = 1 - \frac{k}{N} : k \in \mathbf{N}\}$; probability $x \in (0,1) \cap \mathbf{Q}$ is split in a set $D_x = D_x^- \cup D_x^+$, where D_x^-, in particular, contains the set $\{\lambda = x - \frac{k}{N} : k \in \mathbf{N}\}$ and D_x^+, in particular, contains the set $\{\lambda = x + \frac{k}{N} : k \in \mathbf{N}\}$.

3. Negative probabilities and p-adic ensemble probabilities. Let us consider Example 2.2 of Chapter 3 from the p-adic viewpoint. The series $|S| = 1+2+\ldots+2^k+\ldots = -1$ converges in $\mathbf{Q}_2$. Thus the statistical ensemble S of Example 2.2 has the 2-adic maximal volume -1. Probabilities $\mathbf{p}_k = |S(\lambda = \lambda_k)|/|S| = -2^k$ are infinitely small probabilities. p-adic approach implies that the distribution of quantum systems regarding to values $\lambda = \lambda_j$ of hidden variables has the 2-adic hierarchical structure. The ensemble S has the form of a tower in that the jth floor is 'populated' by quantum systems s with the property $\lambda = \lambda_j$. If we assume that a preparation procedure $\mathcal{E}$ produces portions of quantum systems in the accordance to this tower structure, then there will be extremely unstable behaviour of properties $\lambda = \lambda_j$ in quantum

data which will be used in an experiment (compare with [67]).

The summation in the formula of total probability

$$\mathbf{P}_S(A) = \sum_{k=0}^{\infty} \mathbf{p}_k \sum_{j \in j(A)} \mathbf{p}_{kj} \tag{3.16}$$

is meaningful from 2-adic viewpoint for conditional probabilities $\mathbf{p}_{kj}$ which do not depend on k (for finite sets A).

We now consider Example 2.3 of Chapter 3. Here conditional probabilities $\mathbf{p}_{kl} = -2^l$ are well defined in $\mathbf{Q}_2$. These are infinitely small probabilities. The summation in (3.16) is meaningful. For example, for $A = \{\lambda'_0, ..., \lambda'_{2k}, ...\}$ we have

$$\mathbf{P}_S(A) = (\sum_{l=0}^{\infty} -2^l)(\sum_{j=0}^{\infty} -2^{2j}) = \frac{1}{3}\,.$$

$$\mathbf{P}_S(\bar{A}) = (\sum_{l=0}^{\infty} -2^l)(\sum_{j=0}^{\infty} -2^{2j+1}) = \frac{2}{3}\,.$$

All above series converge in $\mathbf{Q}_2$.

Finally we consider Example 2.4 of Chapter 3. By equality (1.10) the factorial series $\sum_{k=0}^{\infty} k!$ converges in each field $\mathbf{Q}_p$. Thus conditional probabilities

$$\mathbf{p}_{kl} = \frac{l!}{\sum_{k=0}^{\infty} k!}$$

are well defined in each $\mathbf{Q}_p$.

4 Measures

Let X be an arbitrary set and let $\mathcal{R}$ be a ring of subsets of X. The pair $(X, \mathcal{R})$ is called a *measurable space.* The ring $\mathcal{R}$ is said to be *separating* if for every two distinct elements, x and y, of X there exists an $A \in \mathcal{R}$ such that $x \in A, y \notin A$. We shall consider measurable spaces only over separating rings which cover X.

Every ring $\mathcal{R}$ can be used as a base for the zero-dimensional topology[8] which we shall call the $\mathcal{R}$-topology. This topology is Hausdorff iff $\mathcal{R}$ is separating.

[8] A topological space $(X; \tau)$ is zero dimensional if each point $x \in X$ has a basis of clopen (i.e., at the same time open and closed) neighborhoods.

Throughout this section, $\mathcal{R}$ is a separating covering ring of a set X.

A subcollection $\mathcal{S}$ of $\mathcal{R}$ is said to be *shrinking* if the intersection of any two elements of $\mathcal{S}$ contains an element of $\mathcal{S}$. If $\mathcal{S}$ is shrinking, and if f is a map $\mathcal{R} \to K$ or $\mathcal{R} \to \mathbf{R}$, we say that $\lim_{A \in \mathcal{S}} f(A) = 0$ if for every $\epsilon > 0$, there exists an $A_0 \in \mathcal{S}$ such that $|f(A)| \leq \epsilon$ for all $A \in \mathcal{S}, A \subset A_0$.

Let K be a non-Archimedean field with the valuation $|\cdot|_K$.

A *measure* on $\mathcal{R}$ is a map $\mu : \mathcal{R} \to K$ with the properties:

(i) μ is additive;

(ii) for all $A \in \mathcal{R}$, $\|A\|_\mu = \sup\{|\mu(B)|_K : B \in \mathcal{R}, B \subset A\} < \infty$;

(iii) if $\mathcal{S} \subset \mathcal{R}$ is shrinking and has empty intersection, then $\lim_{A \in \mathcal{S}} \mu(A) = 0$.

We call these conditions respectively *additivity*, *bounded*, *continuity*. The latter condition is equivalent to the following: $\lim_{A \in \mathcal{S}} \|A\|_\mu = 0$ for every shrinking collection $\mathcal{S}$ with empty intersection.

Condition (iii) is the replacement for σ-additivity. Clearly (iii) implies σ-additivity. Moreover, we shall see that for the most interesting cases (iii) is equivalent to σ-additivity. Of course, we could in principle restrict our attention to these cases and use the standard condition of σ-additivity. However, in that case we should use some topological restriction on the space X. This implies that we must consider some topological structure on a p-adic probability space. We do not like to do this. We shall develop the theory of p-adic probability measures in the same way as A.N. Kolmogorov(1933) developed the theory of real valued probability measures by starting with an arbitrary set algebra.

Further, we shall briefly discuss the main properties of measures, see [104] for the details. As in Chapter 1, for any set D, we denote its characteristic function by the symbol I_D. For $f : X \to K$ and $\phi : X \to [0, \infty)$, put

$$\|f\|_\phi = \sup_{x \in X} |f(x)|_K \phi(x).$$

We set

$$N_\mu(x) = \inf_{U \in \mathcal{R}, x \in U} \|U\|_\mu$$

for $x \in X$. Then $\|A\|_\mu = \|I_A\|_{N_\mu}$ for any $A \in \mathcal{R}$. We set $\|f\|_\mu = \|f\|_{N_\mu}$.

A *step function* (or $\mathcal{R}$-step function) is a function $f : X \to K$ of the form $f(x) = \sum_{k=1}^N c_k I_{A_k}(x)$ where $c_k \in K$ and $A_k \in \mathcal{R}, A_k \cap A_l = \emptyset, k \neq l$. We set for such a function

$$\int_X f(x)\mu(dx) = \sum_{k=1}^N c_k \mu(A_k).$$

Denote the space of all step functions by the symbol $S(X)$. The integral $f \to \int_X f(x)\mu(dx)$ is the linear functional on $S(X)$ which satisfies the inequality

$$|\int_X f(x)\mu(dx)|_K \leq \|f\|_\mu. \tag{4.1}$$

A function $f : X \to K$ is called *μ-integrable* if there exists a sequence of step functions $\{f_n\}$ such that $\lim_{n\to\infty} \|f - f_n\|_\mu = 0$. The μ-integrable functions form a vector space $L_1(X, \mu)$ (and $S(X) \subset L_1(X, \mu)$). The integral is extended from $S(X)$ on $L_1(X, \mu)$ by continuity. The inequality (4.1) holds for $f \in L_1(X, \mu)$.

Let $\mathcal{R}_\mu = \{A : A \subset X, I_A \in L_1(X, \mu)\}$. This is a ring. Elements of this ring are called μ-measurable sets. By setting $\mu(A) = \int_X I_A(x)\mu(dx)$ the measure μ is extended to a measure on $\mathcal{R}_\mu$. This is the *maximal extension* of μ, i.e., if we repeat the previous procedure starting with the ring $\mathcal{R}_\mu$, we will obtain this ring again.

Set $X_\epsilon = \{x \in X : N_\mu(x) \geq \epsilon\}, X_0 = \{x \in X : N_\mu(x) = 0\}, X_+ = X \setminus X_0$. Every $A \subset X_0$ belongs to $\mathcal{R}_\mu$. We call such sets *μ-negligible.*

Now we construct product measures. Let $\mu_j, j = 1, 2, ..., n$, be measures on (separating) rings $\mathcal{R}_j$ of subsets of sets X_j. The finite unions of the sets $A_1 \times \cdots \times A_n, A_j \in \mathcal{R}_j$, form a (separating) ring $\mathcal{R}_1 \times \cdots \times \mathcal{R}_n$ of $X_1 \times \cdots \times X_n$. Then there exists a unique measure $\mu_1 \times \cdots \times \mu_n$ on $\mathcal{R}_1 \times \cdots \times \mathcal{R}_n$ such that $\mu_1 \times \cdots \times \mu_n(A_1 \times \cdots \times A_n) = \mu_1(A_1) \times \cdots \times \mu_n(A_n)$. We have

$$N_{\mu_1 \times \cdots \times \mu_n}(x_1, ..., x_n) = N_{\mu_1}(x_1) \times \cdots \times N_{\mu_n}(x_n).$$

Let X be a zero-dimensional topological space[9]. We denote the ring of *clopen* (i.e., at the same time open and closed) subsets of X by the symbol $B(X)$ (in fact, this is an algebra). We denote the space of continuous bounded functions $f : X \to K$ by the symbol $C_b(X)$. We use the norm $\|f\|_\infty = \sup_{x\in X} |f(x)|_K$ on this space.

First we remark that *if X is compact and $\mathcal{R} = B(X)$ then the condition (iii) in the definition of a measure is redundant.* If X is not compact then there exist bounded additive set functions which are not continuous.

Let X be zero-dimensional $\mathbf{N}$-compact topological space, i.e., there exists a set S such that X is homeomorphic to a closed subset of $\mathbf{N}^S$. We remark that every product of $\mathbf{N}$-compact spaces is $\mathbf{N}$-compact; every closed subspace of an $\mathbf{N}$-compact space is $\mathbf{N}$-compact. Then every bounded σ-additive

[9]We consider only Hausdorff spaces.

function $\mu : B(X) \to K$ is a measure. On the other hand, if X is a zero-dimensional space such that every bounded σ-additive function $B(X) \to K$ is a measure, then X is $\mathbf{N}$-compact.

In the theory of integration a crucial role is played by the $\mathcal{R}_\mu$-topology, i.e., the (zero-dimensional) topology that has $\mathcal{R}_\mu$ as a base. Of course, $\mathcal{R}_\mu$-topology is stronger that $\mathcal{R}$-topology. Every μ-negligible set is $\mathcal{R}_\mu$-clopen. The following two theorems [104] will be important for our considerations.

Theorem 4.1. (i) *If μ is a measure on $\mathcal{R}$, then N_μ is $\mathcal{R}$-upper semicontinuous, (hence, $\mathcal{R}_\mu$-upper semicontinuous) and for every $A \in \mathcal{R}_\mu$ and $\epsilon > 0$ the set $A_\epsilon = A \cap X_\epsilon$ is $\mathcal{R}_\mu$-compact.*

(ii) Conversely, let $\mu : \mathcal{R} \to K$ be additive. Assume that there exists an $\mathcal{R}$-upper semicontinuos $\phi : X \to [0, \infty)$ such that $|\mu(A)|_K \leq \sup_{x\in A} \phi(x), A \in \mathcal{R}$, and $\{x \in A : \phi(x) \geq \epsilon\}$ is $\mathcal{R}$-compact $(A \in \mathcal{R}, \epsilon > 0)$. Then μ is a measure and $N_\mu \leq \phi$.

Theorem 4.2. *Let $\mu : \mathcal{R} \to K$ be a measure. A function $f : X \to K$ is μ-integrable iff it has the following two properties: (1) f is $\mathcal{R}_\mu$-continuous; (2) for every $\epsilon > 0$, the set $\{x : |f(x)|_K N_\mu(x) \geq \epsilon\}$ is $\mathcal{R}_\mu$-compact.*

We shall also use the following fact.

Theorem 4.3. *Let $f \in L_1(X, \mu)$ and let*

$$\int_A f(x)\mu(dx) = 0 \ \textit{for every} \ A \in \mathcal{R}. \tag{4.2}$$

Then supp $f \subset X_0$.

Proof. Let us assume that f satisfies (4.2) and there exists $x_0 \in X_+$ (hence $N_\mu(x_0) = \alpha > 0$) such that $|f(x_0)|_K = c > 0$. Let $\{f_k\}$ be a sequence of $\mathcal{R}$-step functions which approximates f. For every $\epsilon > 0$ there exist N_ϵ such that $\|f - f_k\|_\mu < \alpha\epsilon$ for all $k \geq N_\epsilon$. In particular, this implies that $|f_k(x_0)|_K \geq c - \epsilon,\ k \geq N_\epsilon$. Then we have

$$\Delta_{B,k} = |\int_B f_k(x)\mu(dx)|_K = |\int_B f_k(x)\mu(dx) - \int_B f(x)\mu(dx)|_K < \alpha\epsilon, \ \ B \in \mathcal{R}.$$

Let

$$f_k(x) = \sum_j c_{kj} I_{B_{kj}}(x), c_{kj} \in K, B_{kj} \in \mathcal{R}, B_{kj} \cap B_{ki} = \emptyset, i \neq j,$$

and let $x_0 \in B_{kj_0}$. If $B \subset B_{kj_0}, B \in \mathcal{R}$, then $\Delta_{B,k} = |c_{kj}|_K |\mu(B)|_K = |f_k(x_0)|_K |\mu(B)|_K < \alpha\epsilon$. On the other hand, as $\|B_{kj_0}\|_\mu \geq \alpha$, then for every

$\delta > 0$, there exists $B \subset B_{kj_0}, B \in \mathcal{R}$, such that $|\mu(B)|_K \geq (\alpha - \delta)$. Thus we obtain for this B: $\Delta_{B,k} \geq (\alpha - \delta)(c - \epsilon)$. By choosing $\epsilon > 0, \delta > 0$, such that $(\alpha - \delta)(c - \epsilon) > \alpha\epsilon$ arrive to the contradiction. ■

We shall use the following simple fact.

Lemma 4.1. *Let $(X_j, \mathcal{R}_j), j = 1, 2$, be measurable spaces and let $f : X_1 \to X_2$ be measurable. If $\mathcal{S}$ is shrinking in $\mathcal{R}_2$ then $f^{-1}(\mathcal{S})$ is shrinking in $\mathcal{R}_1$. If $\mathcal{S}$ has empty intersection, then $f^{-1}(\mathcal{S})$ has also empty intersection.*

Lemma 4.2. *Let $(X_j, \mathcal{R}_j), j = 1, 2$, be measurable spaces and let $\eta : X_1 \to X_2$ be a measurable function. Then, for every measure $\mu : \mathcal{R}_1 \to K$, the function $\mu_\eta : \mathcal{R}_2 \to K$ defined by the equality $\mu_\eta(A) = \mu(\eta^{-1}(A))$ is a measure on $\mathcal{R}_2$ and, for every $\mathcal{R}_2$-continuous function, $h : X_2 \to K$ the following inequality holds:*

$$\|h\|_{\mu_\eta} \leq \|h \circ \eta\|_\mu. \tag{4.3}$$

Proof. We have for every $A \in \mathcal{R}_2$,

$$\|A\|_{\mu_\eta} = \sup\{|\mu(\eta^{-1}(B)) : B \in \mathcal{R}_2, B \subset A\} \leq \|\eta^{-1}(A)\|_\mu < \infty. \tag{4.4}$$

Thus μ_η is bounded. We now prove that μ_η is continuous on $\mathcal{R}_2$. Let $\mathcal{S}$ be shrinking in $\mathcal{R}_2$ which has the empty intersection. By Lemma 4.1 $\eta^{-1}(\mathcal{S})$ is shrinking in $\mathcal{R}_1$ which has also the empty intersection. By (4.4) we obtain that $\lim_{A \in \mathcal{S}} \|A\|_{\mu_\eta} = 0$.

We prove inequality (4.3). Let $h : X_2 \to K$ be $\mathcal{R}_2$-continuous. We wish to prove that $|h(b)|_K N_{\mu_\eta}(b) \leq \|h \circ \eta\|_\mu$ for all $b \in X_2$. So we choose $b \in X_2$ with $h(b) \neq 0$. Then the set $C_b = \{y \in X_2 : |h(y)|_K = |h(b)|_K\}$ is $\mathcal{R}_2$-open. Hence there is a $B \in \mathcal{R}_2$ with $b \in B \subset C_b$. Then

$$|h(b)|_K N_{\mu_\eta}(b) \leq |h(b)|_K \|B\|_{\mu_\eta} \leq |h(b)|_K \|\eta^{-1}(B)\|_\mu =$$

$$\sup_{x \in \eta^{-1}(B)} |h(b)|_K N_\mu(x) \leq \sup_{x \in \eta^{-1}(B)} |(h \circ \eta)(x)|_K N_\mu(x) \leq \|h \circ \eta\|_\mu.$$

■

The following theorem on the change of variables will be important in our probabilistic considerations.

Theorem 4.4. (Khrennikov – van Rooij) *Let $(X_j, \mathcal{R}_j), j = 1, 2$, be measurable spaces and let $\eta : X_1 \to X_2$ be a measurable function, and let $\mu : \mathcal{R}_1 \to K$ be a measure. If $f : X_2 \to K$ is an $\mathcal{R}_2$-continuous function such that the function $f \circ \eta$ belongs to $L_1(X_1, \mu)$, then $f \in L_1(X_2, \mu_\eta)$ and*

$$\int_{X_1} f(\eta(x))\mu(dx) = \int_{X_2} f(y)\mu_\eta(dy).$$

Proof. It suffices to prove that for every $\epsilon > 0$ there exists a $\mathcal{R}_2$-step function g such that $\|f-g\|_{\mu_\eta} \leq \epsilon$ and $\|f \circ \eta - g \circ \eta\|_\mu \leq \epsilon$. By (4.3) the first follows from the second. So we fix $\epsilon > 0$.

By Theorem 4.2 the set

$$A = \{x \in X_1 : |(f \circ \eta)(x)|_K N_\mu(x) \geq \epsilon\}$$

is $\mathcal{R}_1$-compact and therefore contained in an element of $\mathcal{R}_1$. But N_μ is bounded on every element of $\mathcal{R}_1$, so N_μ is bounded on A. We choose $\delta > 0$ so that

$$\delta N_\mu(x) \leq \epsilon \text{ for all } x \in A.$$

As A is compact, $f(\eta(A))$ is also compact. We can cover $f(\eta(A))$ by disjoint closed balls of radius δ : $f(\eta(A)) \subset U_\delta(\alpha_0) \cup ... \cup U_\delta(\alpha_N)$, where α_0 is chosen to be 0 in order to obtain:

$$|\alpha_n|_K \leq |t|_K \text{ for } t \in U_\delta(\alpha_n), n = 0, 1, ..., N. \tag{4.5}$$

For each n, $\mathcal{C}_n = \{C \in \mathcal{R}_2 : C \subset f^{-1}(U_\delta(\alpha_n))\}$ is a collection of open sets covering the compact set $\eta(A) \cap f^{-1}(U_\delta(\alpha_n))$. Thus, for each n there is a $C_n \in \mathcal{C}_n$ such that $\eta(A) \cap f^{-1}(U_\delta(\alpha_n)) \subset C_n$. We now have

$$C_0, ..., C_N \in \mathcal{R}_2,$$

$$C_n \subset f^{-1}(U_\delta(\alpha_n)), n = 0, 1, ..., N,$$

$$\eta(A) \subset C_0 \cup ... \cup C_N.$$

Put $g(x) = \sum_{n=0}^N \alpha_n I_{C_n}(x)$. Then g is a $\mathcal{R}_2$-step function. We wish to show that, for all $a \in X$,

$$\Delta(a) = |(f \circ \eta)(a) - (g \circ \eta)(a)|_K N_\mu(a) \leq \epsilon.$$

Thus, take $a \in X$:

(1) If $a \in A$, then there is a unique n with $\eta(a) \in C_n$. Then $\Delta(a) = |(f \circ \eta)(a) - \alpha_n|_K N_\mu(a) \leq \delta N_\mu(a) \leq \epsilon$.

(2) If $a \notin A$, but $\eta(a) \in C_n$ for some n, then by (4.5) we obtain that $\Delta(a) = |(f \circ \eta)(a) - \alpha_n|_K N_\mu(a) \leq |(f \circ \eta)(a)|_K N_\mu(a) \leq \epsilon$.

(3) If $a \notin C_0 \cup ... \cup C_N$, then $g(\eta(a)) = 0$. Thus $\Delta(a) = |(f \circ \eta)(a)|_K N_\mu(a) \leq \epsilon$ (as $a \notin A$). ∎

Open problem. To find a condition for functions f which is weaker than continuity, but implies the formula of the change of variables.

Further we shall obtain some properties of measures which are specific for measures defined on algebras.

Throughout this section, $\mathcal{A}$ is a separating algebra of a set X. First we remark that if we start with a measure μ defined on the algebra $\mathcal{A}$ then the system $\mathcal{A}_\mu$ of μ-integrable sets is again an algebra.

Proposition 4.1. *Let $\mu : \mathcal{A} \to K$ be a measure. Then for each $\epsilon > 0$, the set X_ϵ is $\mathcal{A}_\mu$-compact.*

This fact is a consequence of Theorem 4.1.

Proposition 4.2. *Let $\mu : \mathcal{A} \to K$ be a measure. Then the algebra $B(X)$ of $\mathcal{A}_\mu$-clopen sets coincides with the algebra $\mathcal{A}_\mu$.*

Proof. We use Theorem 4.2 and the previous proposition. Let $B \in B(X)$. Then I_B is $\mathcal{A}_\mu$-continuous and $\{x : |I_B(x)|_K N_\mu(x) \geq \epsilon\} = B \cap X_\epsilon$. As B is closed and X_ϵ is compact, $B \cap X_\epsilon$ is compact. Thus $B(X) \subset \mathcal{A}_\mu$. ■

As a consequence of Proposition 4.2, we obtain that $C_b(X) \subset L_1(X, \mu)$ (for the space X endowed with $\mathcal{A}_\mu$-topology) and the following inequality holds:

$$|\int_X f(x)\mu(dx)|_K \leq \|f\|_\infty \|X\|_\mu, \quad f \in C_b(X).$$

Let X be zero dimensional topological space. A measure μ defined on the algebra $B(X)$ of the clopen sets is called a *tight* measure. Thus by Proposition 4.2 every measure $\mu : \mathcal{A} \to K$ is extended to a tight measure on the space X endowed with the $\mathcal{A}_\mu$-topology.

Proposition 4.3. *Let $\mu : \mathcal{A} \to K$ be a measure and let $f \in L_1(X, \mu)$. Then f is $(\mathcal{A}_\mu, B(K))$-measurable.*

Proof. By Theorem 4.2 f is $\mathcal{A}_\mu$-continuous. Thus $f^{-1}(B(K)) \subset B(X)$. But by Proposition 4.2 we have that $\mathcal{A}_\mu = B(X)$. ■

5 p-adic probability space

Let $\mu : \mathcal{A} \to \mathbf{Q}_p$ be a measure defined on a separating algebra $\mathcal{A}$ of subsets of the set Ω which satisfies the normalization condition $\mu(\Omega) = 1$. We set $\mathcal{F} = \mathcal{A}_\mu$ and denote the extension of μ on $\mathcal{F}$ by the symbol $\mathbf{P}$. A triple $(\Omega, \mathcal{F}, \mathbf{P})$ is said to be a p-adic *probability space* (Ω is a *sample space*, $\mathcal{F}$ is an algebra of *events*, $\mathbf{P}$ is a *probability*).

As in general measure theory we set

$$\Omega_\alpha = \{\omega \in \Omega : N_{\mathbf{P}}(\omega) \geq \alpha\}, \alpha > 0, \Omega_+ = \cup_{\alpha>0}\Omega_\alpha, \Omega_0 = \Omega \setminus \Omega_+.$$

If a property Ξ is valid on the subset Ω_+ we say that Ξ is valid a.e. (mod $\mathbf{P}$).

Everywhere below (G, Γ) denotes a measurable space over the algebra Γ. Functions $\xi : \Omega \to G$ which are $(\mathcal{F}, \Gamma)$-measurable are said to be random variables.

Everywhere below Y is a zero dimensional topological space. We consider Y as the measurable space over the algebra $B(Y)$. Every random variable $\xi : \Omega \to Y$ is continuous in the $\mathcal{F}$-topology. In particular, $\mathbf{Q}_p$-valued random variables are $(\mathcal{F}, B(\mathbf{Q}_p))$-measurable functions. If $\xi \in L_1(\Omega, \mathbf{P})$, we introduce an *expectation* of this random variable by setting $\mathbf{E}\xi = \int_\Omega \xi(\omega)\mathbf{P}(d\omega)$. We note that every bounded random variable $\xi : \Omega \to \mathbf{Q}_p$ belongs to $L_1(\Omega, \mathbf{P})$.

Let $\eta : \Omega \to G$ be a random variable. The measure $\mathbf{P}_\eta$ is said to be a *distribution* of the random variable. By Theorem 4.4 we have that

$$\mathbf{E}f(\eta) = \int_{\mathbf{Q}_p} f(y)\mathbf{P}_\eta(dy) \tag{5.1}$$

for every Γ-continuous function $f : G \to \mathbf{Q}_p$ such that $f \circ \eta \in L_1(\Omega, \mathbf{P})$. In particular, we have the following result.

Proposition 5.1. *Let $\eta : \Omega \to Y$ be a random variable and let $f \in C_b(Y)$. Then the formula (5.1) holds.*

We shall also use the following technical result.

Proposition 5.2. *Let $\eta : \Omega \to Y$ be a random variable and let $\zeta \in L_1(\Omega, \mathbf{P})$, and let $f \in C_b(Y)$. Then $\xi(\omega) = \zeta(\omega)f(\eta(\omega))$ belongs $L_1(\Omega, \mathbf{P})$ and*

$$\mathbf{E}\xi = \int_{\mathbf{Q}_p \times Y} xf(y)\mathbf{P}_z(dxdy), \quad z(\omega) = (\zeta(\omega), \eta(\omega)).$$

Proof. We have only to show that $\xi \in L_1(\Omega, \mathbf{P})$. This fact is a consequence of Theorem 4.2. ■

The random variables $\xi, \eta : \Omega \to G$ are called independent if

$$\mathbf{P}(\xi \in A, \eta \in B) = \mathbf{P}(\xi \in A)\mathbf{P}(\eta \in B) \text{ for all } A, B \in \Gamma. \tag{5.2}$$

Proposition 5.3. *Let $\xi, \eta : \Omega \to Y$ be independent random variables and functions $f, g \in C_b(Y)$. Then we have:*

$$\mathbf{E}f(\xi)g(\eta) = \mathbf{E}f(\xi)\mathbf{E}g(\eta). \tag{5.3}$$

Proof. If f and g are locally constant functions then (5.3) is a consequence of (5.2). Arbitrary functions $f, g \in C_b(Y)$ can be approximated by

locally constant functions (with the convergence of corresponding integrals) by using the technique developed in the proof of Theorem 4.4. ■

Remark 5.1. In fact, the formula (5.3) is valid for the continuous f, g such that the random variables $f(\xi), g(\eta)$ and $f(\xi)g(\eta)$ belong $L_1(\Omega, \mathbf{P})$.

Proposition 5.4. *Let ξ and η be independent random variables. Then the random vector $z = (\xi, \eta)$ has the probability distribution $\mathbf{P}_z = \mathbf{P}_\eta \times \mathbf{P}_\xi$.*

This fact is the direct consequence of (5.2).

Let ξ and η be respectively $\mathbf{Q}_p$ and G valued random variables and $\xi \in L_1(\Omega, \mathbf{P})$. A *conditional expectation* $\mathbf{E}[\xi|\eta = y]$ is defined as a function $m \in L_1(G, \mathbf{P}_\eta)$ such that

$$\int_{\{\omega\in\Omega:\eta(\omega)\in B\}} \xi(\omega)\mathbf{P}(d\omega) = \int_B m(y)\mathbf{P}_\eta(dy) \text{ for every } B \in \Gamma.$$

Proposition 5.5. *The conditional expectation is defined uniquely a.e.* mod $\mathbf{P}_\eta$.

Proof. We assume that there exist two conditional expectations $m_j \in L_1(G, \mathbf{P}_\eta)$ and $m_1(x_0) \neq m_2(x_0)$ at some point x_0 and $N_{\mathbf{P}_\eta}(x_0) > 0$. Set $m(x) = m_1(x) - m_2(x)$. We have : $\int_B m(x)\mathbf{P}_\eta(dx) = 0$ for every $B \in \Gamma$. To obtain the contradiction, it is sufficient to use Theorem 4.3. ■

As there is no analogue of the Radon-Nikodym theorem in the non-Archimedean case [104], it may happens that a conditional expectation does not exist. Everywhere below we assume that $m(y) = \mathbf{E}[\xi|\eta = y]$ is well defined and moreover, that it belongs to the class $C_b(Y)$.

Proposition 5.6. *Let $\xi : \Omega \to \mathbf{Q}_p, \eta : \Omega \to Y$ be random variables, and $\xi \in L_1(\Omega, \mathbf{P})$. The equality*

$$\mathbf{E}f(\eta)\xi = \mathbf{E}f(\eta(\omega))\mathbf{E}[\xi(\omega)|\eta = \eta(\omega)]$$

holds for every function $f \in C_b(Y)$.

Proof. By Proposition 5.2 we obtain $\mathbf{E}\xi f(\eta) = \int_{\mathbf{Q}_p\times Y} xf(y)\mathbf{P}_z(dxdy)$, where $z(\omega) = (\xi(\omega), \eta(\omega))$. Set for $A \in B(Y)$,

$$\lambda(A) = \int_{\mathbf{Q}_p\times Y} xI_A(y)\mathbf{P}_z(dxdy).$$

As $\lambda(A) = \int_{\eta^{-1}(A)} \xi(\omega)\mathbf{P}(d\omega) = \int_Y m(y)\mathbf{P}_\eta(dy)$, it is a tight measure on Y. Then

$$\int_{\mathbf{Q}_p\times Y} xf(y)\mathbf{P}_z(dxdy) = \int_Y f(y)\lambda(dy) = \int_Y f(y)m(y)\mathbf{P}_\eta(dy) = Ef(\eta)m(\eta).$$

■

Chapter 5

Information and Probability

The title of this chapter may induce the impression that we would discuss the well known connection between probability and information based on entropy. However, we shall discuss ideas which extremely differ from Shannon's ideas that by specifying probabilities of various states it is possible to obtain a quantative measure of information. In the opposite to Shannon's approach we shall demonstrate that, in fact, information induces probability. In particular, we reduce subjective probability to the ensemble probability (proportion) on the space of human ideas. This viewpoint clarifies the use subjective probabilities and implies some consequences for cognitive sciences (in particular, for psychology).

1 Information reality

Our information-probabilistic considerations are based on a new model of reality. In our model physical reality is reality of information. Reality of 17-20 th centuries, *Newton's reality* (or *material reality*), is just a part of general *information reality*. Newton's model of reality gives the description of material systems and motions of such systems (as well as fields associated with material systems). From our point of view material systems give only a particular class of information systems, *transformers of information*. Besides material transformers of information, there exist purely information systems and processes. In particular, the phenomenon of conscious could not be understand on the basis of Newton's model of reality. We propose a model of purely information reality and develop analogues of classical (Newton's)

mechanics on information spaces (in particular, *spaces of ideas* for cognitive systems).

The mathematical basis of Newton's classical mechanics (as well as Einstein's relativity) is given by the system of real numbers **R**. [1] This system describes well dynamics of material systems. However, it is not so useful for studying of information processes and, in particular, cognitive processes. In fact, the absence of adequate mathematical models for cognitive information processes is a consequence of the use of real analysis. It seems that real analysis could not provide the adequate description of cognitive processes. By studying configurations of exited neurons in a domain D_B (called a brain) of real space $\mathbf{R}^3$ we would never understand cognitive information processes and, in particular, the *phenomenon of consciousness.* We have to change a number system and use non-real number systems which provide more adequate description of cognitive information processes. In fact, since the creation of quantum mechanics, it was clear that the mathematical basis of physics must be changed. However, Newton's mathematical model of space was incorporated into a new (quantum) physics. Moreover, the domain of Newton's model was extended to include information phenomena. On one hand, such a situation can be explained by the great positive experience of the use of real numbers in physics and technique. On the other hand, it is a consequence of scientific psychology. It seems that Newton's model of material reality (and, in particular, real analysis) became a kind of scientific religion. Only phenomena described by this model are recognized as physical phenomena.

The ability to form associations is one of main features of information systems. To describe dynamics in information spaces (and, in particular, spaces of ideas), it is natural to use number systems which describe the ability to form associations. One of such number systems, a system of m-adic numbers. From the information viewpoint the ring of m-adic integers $\mathbf{Z}_m$ can be constructed in the following way. Let $A_m = \{0, 1, ..., m-1\}$, where $m \in \mathbf{N}, m > 1$, be an alphabet. We consider infinite information strings with respect to this alphabet

$$x = (\alpha_1, \alpha_2, ..., \alpha_N, ...), \alpha_j \in A_m.$$

We introduce the following metric on the space of information strings. Let

[1] We think that the transition from the Euclidean real space to real Riemannian manifolds does not change very much Newton's model of material reality

$x = (\alpha_j)$ and $y = (\beta_j)$. Then we set

$$\rho_m(x, y) = \frac{1}{m^k} \text{ if } \alpha_j = \beta_j, j = 0, ..., k-1, \text{ and } \alpha_k \neq \beta_k.$$

This metric describes the nearness in the sense of associations of information strings x and y : two information strings (ideas) are close if their first (and the most important) information registers coincide.

Each information string $x = (\alpha_j)_{j=0}^{\infty}$ is identified with a number

$$x = \sum_{j=0}^{\infty} \alpha_j m^j.$$

Such numbers form a system (ring) of m-adic integers $\mathbf{Z}_m$. Geometrically elements $x \in \mathbf{Z}_m$ can be represented as branches of a homogeneous m-adic tree. Thus we can define addition, subtraction and multiplication of branches (which represent infinite information strings, ideas). As it has been already mentioned in Chapter 4, if $m = p > 1$ is a prime number, then it is possible to define division.

In fact, it is possible to use more general number systems which geometrically can be realized as trees with nonconstant number of branches leaving each vertex (see [65]).

2 Dynamics on information spaces

Everywhere below the abbreviation 'I' is used for the word information.

We choose the space $X = \mathbf{Z}_p$ (or multidimensional spaces $X = \mathbf{Z}_p^N$) for the description of information. The X is said to be *information space*[2].

Objects which 'live' in I-spaces are said to be *transformers of information* (I-transformers). I-transformers are not characterized by localization in information p-adic space (or real space). They are characterized by the ability to receive an external information and transform it in a new information.

Each I-transformer τ has internal clocks. A state of the clocks is described by an I-vector $t \in T = \mathbf{Z}_p$ which is called *information time.* The I-time can

[2]We do not assume that X describes the whole information in the universe. By our philosophy there is no absolute physical space (which is typically identified with the real space). There is also no absolute information space. Different information phenomena can be described by different mathematical models for I-spaces. The p-adic model for I-spaces is the simplest from the mathematical point of view.

have different interpretations in different I-models. If τ is a conscious system then t is (self-recognized) time of the evolution of this system. We can say about psychological time of an individual or about (collective) social time of a group of individuals. In fact, we have not to image t as an ordered sequence of time counts. This is only information with describes evolution of τ. In principle, there is no direct relation between I-time and 'physical' time that is used in the model over the reals.

At each instant $t \in T$ of I-time there is defined a *total information state* (I-state) $q(t) \in X$ of τ. It describes the position of τ in the I-space X. The 'life'-trajectory of τ can be identified with the trajectory $q(t)$ in X.

We use an analogue of the Hamiltonian dynamics on the I-spaces [3]. As usual, we introduce the quantity $p(t) = \dot{q}(t)$ $(= \frac{d}{dt}q(t))$ which is the information analogue of the momentum. However, here we prefer to use a psychological terminology. The quantity $p(t)$ is said to be a *motivation* (for changing of the I-state $q(t)$).

The space $\mathbf{Z}_p \times \mathbf{Z}_p$ of points $z = (q,p)$ where q is the I-state and p is the motivation is said to be a phase I-space. As in the ordinary Hamiltonian formalism, we assume that there exists a function $H(q,p)$ (I-Hamiltonian) on the phase I-space which determines the motion of τ in the phase I-space:

$$\dot{q}(t) = \frac{\partial H}{\partial p}(q(t),p(t)),\ q(t_0) = q_0,\ \ \dot{p}(t) = -\frac{\partial H}{\partial q}(q(t),p(t)),\ p(t_0) = p_0.$$

The I-Hamiltonian $H(p,q)$ has the meaning of an *I-energy.* In principle, I-energy is not related to the usual physical energy.

The simplest I-Hamiltonian $H_f(p) = \alpha p^2$, $\alpha \in Z_p$ describes the motion of a free I-transformation τ, i.e., an I-transformer which uses only self-motivations for changing of its I-state $q(t)$. Here by solving the system of the Hamiltonian equations we obtain: $p(t) = p_0$, $q(t) = q_0 + 2\alpha p_0(t - t_0)$ [4]. The motivation p is the constant of this motion. Thus the free I-transformer "does not like" to change its motivation p_0 in the process of the motion in the I-space. If we change coordinates, $q' = (q - q_0)/k$, $k = 2\alpha p_0$, then we

[3]In fact, this is an application to the I-theory of the Hamiltonian p-adic formalism developed in [107] (and generalized in [55]).

[4]In fact, this simplest I-system is not trivial from the mathematical viewpoint. There exist other solutions which are nonanalytic (but smooth), see [97], [104], [107], [60]. These solutions may also have an interesting I-interpretation. We shall discuss this problem later.

see that the dynamics of the free I-transformer coincides with the dynamics of its I-time.

In the general case the I-energy is the sum of the I-energy of motivations $H_f = \alpha p^2$ (which is an analogue of the kinetic energy) and the potential I-energy $V(q)$:

$$H(q,p) = \alpha p^2 + V(q).$$

The potential $V(q)$ is determined by *fields of information.*

In the Hamiltonian framework we can consider interactions between I-transformers $\tau_1, \dots, \tau_N$. These I-transformers have the I-times $t_1, \dots, t_N$ and I-states $q_1(t_1), \dots, q_N(t_N)$. By our model we can describe interactions between these I-transformers only in the case in that there is the possibility to choose the same I-time t for all of them. In this case we can consider the evolution of the system of the I-transformers $\tau_1, \dots, \tau_N$ as a trajectory in the I-space $\mathbf{Z}_p^N = \mathbf{Z}_p \times \cdots \times \mathbf{Z}_p$, $q(t) = (q_1(t), \dots, q_N(t))$.

We think that the condition of consistency

$$t_1 = t_2 = ... = t_N = t \tag{2.1}$$

plays the crucial role in many psychological experiments. We can not obtain sensible observations for interactions between arbitrary individuals. There must be a process of learning for the group $\tau_1, \dots \tau_N$ which reduces I-times $t_1, \dots, t_N$ to the unique I-time t.

Thus, let us consider a group $\tau_1, \dots, \tau_N$ of I-transformers with the internal time t. The dynamics of I-states and motivations is determined by the I-energy; $H(q,p)$, $q \in \mathbf{Z}_p^N$, $p \in \mathbf{Z}_p^N$. It is natural to assume that

$$H(q,p) = \sum_{j=1}^{N} \alpha_j p_j^2 + V(q_1, \dots, q_N), \ \alpha_j \in \mathbf{Z}_p.$$

Here $H_f(p) = \sum_{j=1}^{N} \alpha_j p_j^2$ is the total energy of motivations for the group $\tau_1, \dots, \tau_N$ and $V(q)$ is the potential energy. As usual, to find a trajectory in the phase I-space $\mathbf{Z}_p^N \times \mathbf{Z}_p^N$, we need to solve the system of Hamiltonian equations: $q_j = \frac{\partial H}{\partial p_j}$, $p_j = -\frac{\partial H}{\partial q_j}$, $q_j(t_0) = q_0$, $p_j(t_0) = p_0$.

Consequences for cognitive and social sciences and psychology:

1. *Energy and information.* In our model transmission of information is determined by the I-energy which is the sum of I-energy of motivations and potential I-energy. In principle, this process need no physical energy.

2. *Distance and information.* I-processes may evaluate in an I-space which differs from the real space (the absolute Newton space or spaces of general relativity). Therefore the real ('physical') distance between I-transformers does not play the crucial role in processes of I-interactions.

3. *Time and information.* I-dynamics is dynamics with respect to I-time t. There may be a correspondence $t_{\text{phys}} = g(t)$ between real time $t_{\text{phys}} \in \mathbf{R}$ and I-time $t \in \mathbf{Z}_p$. This correspondence may not preserve distances. Thus some I-interactions may be interpreted as *I-interactions with the past or future.*

4. *Motivation.* A motion in the I-space depends, not only on the initial I-state q_0, but also on the initial motivation p_0. Moreover, the Hamiltonian structure of the equations of motion implies that the motivation $p(t)$ plays the important role in the process of the evolution. Thus I-dynamics is, in fact, dynamics in phase I-space.

5. *Social phenomena.* By our model any social group G can be described by a system $\tau_1, \ldots, \tau_N$ of coupled I-transformers. There exists an I-potential $V(q_1, \ldots, q_N)$ which determines an I-interaction between members of G. For example, democratic societies are characterized by uniform I-potentials $V = \sum \Phi(q_i - q_j)$. Here a contribution into the potential I-energy does not depend on an individual. On the other hand, hierarchic societies are characterized by I-potentials of the form:

$$V = A_0 \sum_{j \neq 0} \Phi(q_0, q_j) + A_1 \sum_{j \neq 0,1} \Phi(q_1, q_j) + \cdots$$

$$+ A_k \sum_{j \neq 0, \ldots, k} \Phi(q_k, q_j) + B \sum_{i,j \neq 0, \ldots, k} \Phi(q_i, q_j),$$

where $|A_0|_p >> |A_1|_p >> \cdots >> |A_k|_p >> |B|_p$. These potentials describe the hierarchy $\tau_0 \to \tau_1 \to \cdots \to \tau_k \to (\tau_{k+1}, \ldots, \tau_N)$. The I-transformer τ_0 can be a political, national or state leader or a God.

Remark 2.1. (Active information) Our ideas about information and information field are similar to the ideas of D. Bohm and B. Hiley (BH) [13] (see especially p. 35-38). As BH, we do not follow "Shannon's ideas that there is a quantative measure of information that represents the way in which the state of a system is uncertain for us", [13]. We also consider information as an *active information.* Such information interacts with I-transformers. As a consequence of such interactions I-transformers produce new information. The only distinguishing feature is that material objects are not involved in our formalism. According to BH active information interacts with material objects (for example, the ship guided by radio waves). BH assume that information fields have nonzero physical energy which directs other (probably very large) physical energy. However, physical energies are not involved in our model. Thus we need not assume that I-fields

have some physical energy. In particular, we need not try to find (as BH, [13], p.38) an origin of such an energy. We also remark that BH discuss only quantum ψ-fields. We shall use both classical and quantum I-fields.

In fact, BH consider objects which are similar to our I-transformers. However, they think that such objects must always have the material realization (i.e., they must be represented in real space). In our model we need not such a realization. We think that there might be I-transformers (such as the consciousness) which could not be in principle represented in the real space.

BH discussed a difference between 'active' and 'passive' information. In fact, our model supports their conclusion that "all information is at least potentially active and that complete passivity is never more than an abstraction...", [13], p.37. If an I-transformer τ moves in the field of forces ψ (classical or quantum), then the information $x \in \text{supp V}$ is active for τ and the information $x \in \mathbf{Z}_p^m \setminus \text{supp V}$ is passive for τ. Let $v = V(t, x)$ be a time dependent potential. Then the set of active information $X(t) = \text{supp V(t)}$ evolves in I-space. Thus some passive information becomes active and vice versa.

Remark 2.2. (Mind-matter interaction) By our model there is no difference between 'material' and 'nonmaterial' I-transformers. Interactions between mind and matter are only particular interactions between I-transformers. On one hand, material objects can generate I-fields (fields of I-forces) which change I-states of cognitive objects. On the other hand, cognitive objects can also generate I-fields which change I-states of material objects. Later we shall see that cognitive objects are quantum I-objects. Thus mind-matter interaction is a particular case of interactions between quantum and classical objects.

3 Information velocity, acceleration, mass and force, Newton's laws

We have considered dynamics of I-transformers of the unit mass. There the coefficient v of proportionality between the variation δq of the I-state and the variation δt of I-time t: $\delta q = v\delta t$, was considered as a motivation. In the general case the motivation p may not coincide with v. Let us assume that the motivation p is proportional to v, $p = mv, m \in \mathbf{Z}_p$. This coefficient m of proportionality is called an *I-mass*. We also call v an *I-velocity*. Thus $\delta q = \frac{p}{m}\delta t$.

Let τ_1 and τ_2 be two I-transformers with the I-masses m_1 and m_2 and let $|m_1|_p > |m_2|_p$. Let τ_1 and τ_2 have the variations δt_1, δt_2 of I-time of the

same p-adic magnitude, $|\delta t_1|_p = |\delta t_2|_p$, and let these variations generate the variations δq_1 and δq_2 of their I-states of the same p-adic magnitude, $|\delta q_1|_p = |\delta q_2|_p$. To make such a change of the I-state, τ_1 need larger motivation: $|p_1|_p = |\frac{\delta q}{\delta t}|_p |m_1|_p > |p_2|_p = |\frac{\delta q}{\delta t}|_p |m_2|_p$. Thus the I-mass is a measure of an *inertia of information.* We define a *kinetic* I-energy by $T = \frac{1}{2m}p^2$. A variation δt of I-time t implies also a variation δp of the motivation p: $\delta p = f\delta t$. The coefficient f of proportionality is called an *I-force.* Thus any change of the motivation is due to the action of an I-force f. If $f = 0$ then $\delta p = 0$ for any variation δt of t. Thus an I-transformer cannot change its motivation in the absence of I-forces. By analogue with the usual physics we call the coefficient a of a proportion between the variation δv of the I-velocity v and the variation δt of the I-time t, $\delta v = a\delta t$, an *I-acceleration.* Thus $\delta p = am\delta t$. This relation can be rewritten in the form of an information analogue of the second Newton law:

$$ma = f \text{ or } \dot{p} = f. \tag{3.1}$$

An I-force f is said to be a *potential* force if there exists a function $V(q)$ such that $f = -\frac{\partial V}{\partial q}$ where V is called the potential, or potential energy. The total I-energy H is defined as the sum of the kinetic and the potential I-energies, $H(q,p) = \frac{1}{2m}p^2 + V(q)$. The Hamiltonian equation $\dot{p} = -\frac{\partial H}{\partial q}$ coincides with the Newton equation $\dot{p} = f$.

Example 3.1. (Hooke's I-system). Let the I-force f be proportional to the I-state q, $f = m\beta^2 q$, where m is the I-mass and $\beta \in \mathbf{Z}_p$ is a coefficient of the interaction. Here (3.1) gives the equation $\ddot{q} = \beta^2 q$. As $f = -\frac{\partial V}{\partial q}$, $V(q) = -\frac{m\beta^2}{2}q^2$ and $H(q,p) = \frac{p^2}{2m} - \frac{m\beta^2 q^2}{2}$; the Hamiltonian equations are $\dot{q} = p/m$ and $\dot{p} = m\beta^2 q$. Their solutions have the form $g(t) = ae^{\beta t} + be^{-\beta t}$. The I-state $q(t)$ and motivation $p(t)$ are defined only for instants of I-time which satisfy the inequality

$$|\beta t|_p \le r_p. \tag{3.2}$$

This condition can be considered as a restriction for the magnitude of the I-force. If the coefficient of the interaction $|\beta|_p \le r_p$, then dynamics $q(t)$ of the I-state is well defined for all $t \in \mathbf{Z}_p$. Larger forces imply the restriction condition for I-time. Let $|\beta|_p = 1$. If $p \neq 2$ then (3.2) has the form $t \in U_{1/p}(0)$, i.e., $t = \alpha_1 p + \alpha_2 p^2 + \cdots$. Thus the I-state $q(t)$ of the I-transformer τ can be defined (observed) only for the instants of time $t_0 = 0, t_1 = p, ..., t_{p-1} = (p-1)p, ...$. If $p = 2$ then (3.2) has the form $t \in U_{1/4}(0)$, i.e., and $t = \alpha_2 2^2 + \alpha_3 2^3 + \cdots$. Thus the I-state $q(t)$ of τ can be defined (observed) only for the instants of time $t_0 = 0, t_1 = 4, t_2 = 8, ...$.

Let $f = -m\beta^2 q$, i.e., $V(q) = \frac{m\beta^2 q^2}{2}$ and $\ddot{q} = -\beta^2 q$. Here $q(t)$ and $p(t)$ have the form $g(t) = a\cos\beta t + b\sin\beta t$. Here we also have the restriction relation (3.2). In the contrary to the real case the p-adic trigonometric functions are not periodical. There is no analogue of oscillations for the I-process described by an analogue of Hooke's law.

Let us consider the solution of the Hamiltonian equations with the initial conditions $q(0) = 0$ and $p(0) = m\beta$: $q(t) = \sin\beta t$, $p(t) = m\beta\cos\beta t$. We have $qp = (m\beta/2)\sin 2\beta t$. By using the p-adic equality $|\sin a|_p = |a|_p$ we get $|qp|_p = |m\beta|_p|\beta t|_p$. Relation (3.2) implies

$$|q|_p|p|_p \leq |m\beta|_p r_p. \tag{3.3}$$

This is a *restriction relation* for the trajectory $(q(t), p(t))$ in the phase I-space. Let $\beta = 1/m$. Then (3.3) gives $|q|_p|p|_p \leq r_p$. If the motivation p is strong $|p|_p = 1$, then q can be only of the form $q = \alpha_1 p + \alpha_2 p^2 + \cdots$, $p \neq 2$ and $q = \alpha_2 2^2 + \alpha_3 2^3 + \cdots$, $p = 2$. If the motivation p is rather weak then the I-state q of an I-transformer can be arbitrary.

We discuss the role of the I-mass in the restriction relation (3.3). There the decrease of the I-mass implies more rigid restrictions for possible I-states (for the fixed magnitude of the motivation). [5].

The restriction relation (3.3) is an analogue of the Heisenberg uncertainty relations in the ordinary quantum mechanics. However, we consider a classical (i.e., not quantized) I-system. Therefore a classical I-system can have behaviour that is similar to quantum behaviour.

4 Mathematical 'pathologies'

In p-adic analysis the condition $f \equiv 0$ does not imply that a differentiable function f is a constant, see [97]. There exist complicated continuous motions $(q(t), p(t))$ in the I-phase space for I-transformers with zero I-energy ($\dot{q} \equiv 0$ or $\dot{p} \equiv 0$).

In psychological models these motions can be interpreted as motions without any motivation. Such motions need no information force. On the other hand, we can consider an I-potential $V(q)$ such that $\frac{\partial V}{\partial q} = 0$. Here the potential I-energy $V(q)$ can have complicated behaviour on the I-space $X = \mathbf{Z}_p$.

[5] In psychological (social) applications we get that the individual (or a group of individuals) with a small magnitude of I-mass and the strong motivations will have quite restricted set of I-states.

At the same time the I-force $f = 0$. Thus there may exist I-fields which do not induce any I-force.

All mathematical pathologies can be eliminated by the consideration of analytical functions. If $f \equiv 0$ and f is analytic then $f =$ constant[6].

5 Quantum mechanics on information spaces

It is quite natural to quantize classical mechanics on information spaces over $\mathbf{Z}_p$. We give the following reasons for such quantization. Observations over I-quantities are statistical observations. We have to study statistical ensembles of I-transformers (instead studying of an individual I-transformer). Such statistical ensembles are described by quantum states ϕ. As usual in quantum formalism, we can assume that a value λ of an I-quantity A can be measured in the state ϕ with some probability $\mathbf{P}_\phi(A = \lambda)$. This ideology is nothing than the application of the *statistical (ensemble) interpretation of quantum mechanics* to the information theory. By this interpretation any measurement process has two steps: (1) a preparation procedure $\mathcal{E}$; (2) a measurement of a quantity B in the states ϕ which were prepared with the aid of $\mathcal{E}$.

Let us consider these steps in the information framework. By $\mathcal{E}$ we have to select a statistical ensemble ϕ of I-transformers on the basis of some I-characteristics. Typically in quantum physics a preparation procedure $\mathcal{E}$ is realized as a filter based on some physical quantity A, i.e., we select elements which satisfy the condition $A = \mu$ where μ is one of the values of A. We can do the same in quantum I-theory. An I-quantity A is chosen as a filter, i.e., I-transformers for the statistical ensemble ϕ are selected by the condition $A = \mu$ where $\mu \in \mathbf{Z}_p$ is some information. For example, we can choose $A = p$, the motivation, and select the statistical ensemble $\phi = \phi(p = \mu)$ of I-transformers which have the same motivation $\mu \in \mathbf{Z}_p$. Then we realize

[6]In psychological models we can interpret analytical trajectories in the phase I-space as normal behaviour, i.e., an individual need a motivation for the change of a psychological state. Here we can observe some psychological (information) force which induces this change. There is a psychological (information) field which generates this force. Trajectories (non-analytical) with zero motivation are interpreted as abnormal psychological behaviour (probably such trajectories correspond to mental diseases; on the other hand, they may explain anomalous phenomena). Here an individual changes his psychological state without any motivation in the absence of any psychological (information) force. Here, in fact, the p-adic generalization of the Hamiltonian formalism does not work. We need to propose a new physical formalism to describe such phenomena.

the second step of a measurement process and measure some information quantity B in the state $\phi_{(p=\mu)}$. For example, we can measure the I-state q of I-transformers belonging to the statistical ensemble described by $\phi_{(p=\mu)}$. We obtain the probability distribution $\mathbf{P}(q = \lambda | p = \mu)$, $\lambda, \mu \in \mathbf{Z}_p$ (a probability that I-transformer has the I-state $q = \lambda$ under the condition that it has the motivation $p = \mu$). It is also possible to measure the I-energy E of I-transformers. We obtain the probability distribution $\mathbf{P}(E = \lambda | p = \mu)$, $\lambda, \mu \in \mathbf{Z}_p$. On the other hand, we can prepare the statistical ensemble $\phi_{(q=\mu)}$ by fixing some information $\mu \in \mathbf{Z}_p$ and selecting all I-transformers which have the I-state $q = \mu$. Then we can measure motivations of these I-transformers and we obtain a probability distribution $\mathbf{P}(p = \lambda | q = \mu)$.

Other possibility is to use a generalization of the individual interpretation of quantum mechanics. By this interpretation a wave function $\psi(x), x \in \mathbf{R}^n$, describes the state of an individual quantum particle. In the same way we may assume that a wave function $\psi(x), x \in \mathbf{Z}_p^n$, on the I-space describes the state of an individual I-transformer τ [7].

In fact, a mathematical model for quantum I-formalism has been already constructed. This is quantum mechanics with p-adic valued functions, see [57], [60], [4]. We present briefly this model. The space of quantum states is represented as a p-adic Hilbert space $\mathcal{K}$ (see [55], [60], [65] for the theory of such spaces). This is a $\mathbf{Q_p}$-linear space which is a Banach space (with the norm $\|\cdot\|$) and on which is defined a symmetric bilinear form $(\cdot,\cdot)$: $\mathcal{K} \times \mathcal{K} \to \mathbf{Q_p}$. This form is called an inner product on $\mathcal{K}$. It is assumed that the norm and the inner product are connected by the Cauchy-Bunaykovski-Schwarz inequality: $|(x,y)|_p \leq \|x\| \|y\|, \ x, y \in \mathcal{K}$. By definition quantum I-state ϕ is an element of $\mathcal{K}$ such that $(\phi, \phi) = 1$; quantum I-quantity A is a symmetric bounded operator $A : \mathcal{K} \to \mathcal{K}$, i.e., $(Ax, y) = (x, Ay), \ x, y \in \mathcal{K}$[8]. We discuss a statistical interpretation of quantum states in the case of a discrete spectrum of A. Let $\{\lambda_1, \ldots, \lambda_n, \ldots\}$, $\lambda_j \in \mathbf{Z}_p$ be eigenvalues of A, $A\phi_n = \lambda_n \phi_n$, $\phi_n \in \mathcal{K}$, $(\phi_n, \phi_n) = 1$. The eigenstates ϕ_n of A are considered as pure quantum I-states for A, i.e., if the system of I-transformers is described

---

[7] As we have seen, the problem of interpretations is the important problem of ordinary quantum mechanics on real space. The same problem arises immediately in our quantum I-theory. We do not like to start our investigation with a hard discussion on the right interpretation. We can be quite pragmatic and use both interpretations by our convenience.

[8] In p-adic models we do not need to consider unbounded operators, because all quantum quantities can be represented by bounded operators, see [4], [65].

by the state ϕ_n then the I-quantity A has the value $\lambda_n \in \mathbf{Z}_p$ with probability 1. Let us consider a mixed state

$$\phi = \sum_{n=1}^{\infty} q_n \phi_n, \;\; q_n \in \mathbf{Q}_p, \tag{5.1}$$

where $(\phi, \phi) = \sum_{n=1}^{\infty} q_n^2 = 1$[9]. By the statistical interpretation of ϕ if we perform a measurement of the I-quantity A for I-transformers belonging to the statistical ensemble described by ϕ then we obtain the value λ_n with probability $P(A = \lambda_n | \phi) = q_n^2$.

The main problem (or the advantage?) of this quantum model is that these probabilities belong to the field of p-adic numbers $\mathbf{Q}_p$. The simplest way is to eliminate this problem by considering only finite mixtures (5.1) for which $q_n \in \mathbf{Q}$ (the field of rational numbers $\mathbf{Q}$ is a subfield of $\mathbf{Q}_p$). In this case the quantities $\mathbf{P}(A = \lambda_n | \phi) = q_n^2$ can be interpreted as usual probabilities (for example, in the framework of Kolmogorov's theory). Therefore we may assume that there exist (can be prepared) quantum I-states ϕ which have the standard statistical interpretation: when the number N of experiments tends to infinity, the frequency $\nu_N(A = \lambda_n | \phi)$ of an observation of the information $\lambda_n \in \mathbf{Z}_p$ tends to the probability q_n^2.

However, we can also use p-adic probabilistic models of Chapter 4. By using the p-adic frequency probability model for the statistical interpretation of quantum I-states we may assume that there exists I-states ϕ (ensembles of I-transformers) such that the relative frequencies $\nu_N(A = \lambda_n | \phi)$ have no limit in $\mathbf{R}$, i.e., we cannot apply the standard law of large numbers in this situation. Hence if we perform measurements of an I-quantity A for such a quantum I-state and study the observed data by using the standard statistical methods (based on real analysis), then we shall not obtain the definite result. There will be only random fluctuations of relative frequencies, see [60] [10].

The evolution of a p-adic wave function is described by an I-analogue of

[9] As in the usual theory of Hilbert spaces, eigenvectors corresponding to different eigenvalues of a symmetric operator are orthogonal.

[10] Such a behaviour can be related to psychological experiments. Here the possibility of the use of p-adic probability models gives the important consequence for scientists doing experiments with a statistical I-data: the absence of the statistical stabilization (random fluctuations) does not imply the absence of I-phenomenon. This statistical behaviour may have the meaning that this I-phenomenon cannot be described by the standard Kolmogorov probability model.

the Schrödinger equation:

$$\frac{h_p}{i}\frac{\partial \psi}{\partial t}(t,x) = \frac{h_p^2}{2m}\frac{\partial^2 \psi}{\partial x^2}(t,x) - V(t,x)\psi(t,x), \tag{5.2}$$

where m is the I-mass of a quantum I-transformer. Here a constant h_p plays the role of the Planck constant. By pure mathematical reasons (related to convergence of p-adic exponential and trigonometric series) it is convenient to choose $h_p = \frac{1}{p}$.

We may also present some physical arguments for such a choice. In ordinary quantum mechanics the Planck constant is related to the measure of discretization. The constant $h_p = \frac{1}{p}$ is related to the level of discretization of information.

We use the factor $i = \sqrt{-1}$ in (5.2), because we like to have the total coincidence with formulas of the ordinary quantum mechanics. As we have already noted, in the p-adic case the functions $e^{i\alpha x}$ and $e^{\alpha x}$ have the same (non-oscillating) behaviour. Therefore, in principle, we can use the analogue of (5.2) in that the factor i is omitted.

The use of i implies the consideration of the extension $\mathbf{Q}_p(i) = \mathbf{Q}_p \times i\mathbf{Q}_p$ of $\mathbf{Q}_p$. Elements of this extension have the form $z = a + ib, a, b \in \mathbf{Q}_p$. This extension is well defined for $p = 3, \mathrm{mod}\ 4$. As usual, we introduce a convolution $\bar{z} = a - ib$; here we have $z\bar{z} = a^2 + b^2$. In what follows we assume that wave functions take values in $\mathbf{Z}_p(i) = \mathbf{Z}_p \times i\mathbf{Z}_p$.

Example 5.1. (A free I-transformer). Let the potential $V = 0$. Then the solution of the Schrödinger equation corresponding to the I-energy $E = \frac{\mathbf{p}^2}{2m}$ has the form[11] :

$$\psi_{\mathbf{p}}(t,x) = e^{i(\mathbf{p}x - Et)/h_p}. \tag{5.3}$$

By the choice $h_p = 1/p$ this function is well defined for all $x \in \mathbf{Z}_p$ and $t \in \mathbf{Z}_p$. As $\psi\bar{\psi} \equiv 1$, this wave function describes the uniform (p-adic probability) distribution, see Chapter 6, on the ring of p-adic integers $\mathbf{Z}_p$. Thus an I-transformer τ in the state ψ can be observed with equal probability in any state $x \in \mathbf{Z}_p$. In this sense behaviour of the free I-transformer is similar to behaviour of the ordinary free quantum particle. On the other hand, there is

[11]We note that formal expressions for analytical solutions of p-adic differential equations coincide with the corresponding expressions in the real case (in fact, we can consider these equations over arbitrary number field, see [60], [55]). However, behaviours of these solutions are different.

no analogue of oscillations [12]: $\psi_{\mathbf{p}}(t,x) = \cos(\mathbf{p}x - Et)/h_p + i\sin(\mathbf{p}x - Et)/h_p$, and $|\cos(\mathbf{p}x - Et)/h_p|_p = 1, |\sin(\mathbf{p}x - Et)/h_p|_p = |(\mathbf{p}x - Et)/h_p|_p$.

We consider a psychological (and social) consequence of Example 5.1: *in the absence of the external potential the same motivation* $\mathbf{p}$ *may imply any I-state* $x \in \mathbf{Z}_p$.

Let us consider mixtures of states of the form (5.3). We set $t = 0$. Let $\psi(x) = a_1\psi_{\mathbf{p}_1} + a_2\psi_{\mathbf{p}_2},\ a_1, a_2 \in \mathbf{Z}_p$.

If we compute $<\psi,\psi> = \int_{\mathbf{Z}_p} \psi(x)\overline{\psi(x)}dx$ (where dx is a uniform p-adic valued distribution on $\mathbf{Z}_p$) we see a large difference with ordinary quantum mechanics: $<\psi,\psi> \neq a_1\bar{a}_1 + a_2\bar{a}_2$. There is nonzero correlation term. For $\alpha = (\mathbf{p}_1 - \mathbf{p}_2)/h_p$, we have [69]:

$$T(\alpha) = <\psi_{\mathbf{p}_1}, \psi_{\mathbf{p}_2}> + <\psi_{\mathbf{p}_2}, \psi_{\mathbf{p}_1}> = \frac{\alpha \sin\alpha}{1 - \cos\alpha}.$$

Thus there are correlations between the motivations $\mathbf{p}_1$ and $\mathbf{p}_2$ in the state ψ.

Example 5.2. (Quantum Hooke's system) To give an example of a Hamiltonian with discrete spectrum, we consider the formal p-adic generalization of the Hamiltonian of a harmonic oscillator:

$$\hat{H} = -\frac{h_p^2}{2m}\frac{d^2}{dx^2} - \frac{1}{2}m\omega^2x^2 - \frac{1}{2},$$

where m is the I-mass. We consider ω simply as the coefficient of interaction (there is no analogue of harmonic oscillations). The operator $\hat{H}$ has eigenvalues $E_n = h_p\omega n, n = 0, 1, ...$ (see [65]). However, in the p-adic case the difference between continuous and discrete spectra is not so strong (for each E_n, we have $E_n = \lim_{k\to\infty} E_{l_k}, l_k \neq n$). On the other hand, discreteness of a spectrum, of course, induces some restrictions on values (information) which can be observed.

[12] Is it possible to reproduce oscillations with respect to ordinary real time on the basis of the information model? It could be done by the time scaling. Let $f : \mathbf{Z}_p \to \mathbf{Z}_p$ be an arbitrary continuous function. Then $f(t + kp^n) \approx f(t)$ for all $k \in \mathbf{Z}$ for sufficiently large n (uniformly for $t \in \mathbf{Z}_p$). Let $t_{\text{phys}} = g(t)$ be a law of the correspondence between I-time $t \in \mathbf{Z}_p$ and real time $t_{\text{phys}} \in \mathbf{R}$. If $2\pi = g(p^n)$ then the p-adic continuity will imply the periodicity in real time. Therefore, the ordinary wave behaviour is nothing other than a consequence of continuity of information flows and the appropriative choice of a time scale. Depending on the time scale an I-process may or may not exhibit wave behaviour in the real picture of reality.

6 The pilot wave theory for cognitive systems

Let us consider a system of N I-transformers, $\tau_1, ..., \tau_N$, with Hamiltonian $H = \sum_k \frac{p_k^2}{2m_j} + \sum_{k>i} V_{ki}(x_k - x_i)$. The wave function $\psi(t,x), x = (x_1, ..., x_N)$, $x_k \in \mathbf{Z}_p^m$ (where m is the dimension of I-space which is used for the description of the I-state of τ_j) evolves according to the Schrödinger equation (5.2). A purely mathematical consequence of this is that

$$\frac{\partial}{\partial t}\,\rho(t) + \sum_k \frac{\partial}{\partial x_k} j_k(x,t) = 0,$$

where $\rho(t,x) = \psi(t,x)\overline{\psi(t,x)}$ is a probability density on the configuration I-space $\mathbf{Z}_p^{mN}$ and

$$j_k(x,t) = m_k^{-1}\mathrm{Im}(\overline{\psi(\mathrm{t},\mathrm{x})}\frac{\partial}{\partial \mathrm{x_k}}\psi(\mathrm{t},\mathrm{x})).$$

As in the ordinary Bohm's formalism, we assume that a quantum I-transformer τ_k has at any I-time[13] well defined I-state x_k and motivation p_k. I-states x_k evolve according to

$$\dot{x}_k(t) = \frac{j_k(t,x)}{\rho(t,x)}. \tag{6.1}$$

It is assumed that the wave function $\psi(t,x)$ drives I-transformers $\tau_1, ..., \tau_N$. If we generalize ideas of J. Bell [12], then $\psi(t,x)$ can be considered as a new information field which is generated by an 'information life' of the system of I-transformers. At the moment we could not provide the clear explanation of the origin of this information field. It is only possible to observe its influence to trajectories of I-transformers in the I-space $\mathbf{Z}_p^m$. By Bohm's ideology we consider a new I-potential (a quantum I-potential) Q generated by $\psi(t,x)$. Then the quantum I-motion can be considered as a perturbation of the classical I-motion (based on the classical potential V).

This model gives the natural description of an evolution of the I-states of a system of cognitive systems $\tau_1, ..., \tau_N$.

We start with an attempt to describe the work of a brain τ in the framework of the pilot-wave I-theory. The τ has an incredibly complex internal structure which generates a new information field given by brain's wave

[13] Of course, we assume that I-times $t_1, ..., t_N$ of the I-transformers $\tau_1, ..., \tau_N$ satisfy the condition of consistency (2.1).

function $\psi(t,x)$ [14]. We claim that the field $\psi(t,x)$ induced by the brain τ is nothing other than a conscious field. Thus conscious processes are quantum I-processes. A conscious (quantum) motion in phase I-space differs from an unconscious (classical) motion. This difference is due to a quantum I-potential Q.

We consider now a system S of brains $\tau_1, ..., \tau_N$. The wave function $\psi(t,x)$ of the S depends on I-states of all brains in the S. Thus motions of these brains in phase I-space $\mathbf{Z}_p^{2m}$ are not independent. At the same time there might be no classical potential V which induces such a dependence. Of course, if (as in ordinary real formalism) $\psi(t,x) = \prod_{j=1}^N \psi_j(t,x)$, then the I-motions of brains τ_j are independent. There are no correlations between consciousness of different brains.

Remark 6.1. (Non-locality) This is the good place to discuss the problem of non-locality of the pilot wave formalism. Some authors (see, for example, [30]) consider non-locality as one of the main difficulties of the pilot wave formalism. However, non-locality is not a difficulty in our pilot I-wave formalism. This is non-locality in the I-space. Such non-locality can be natural for some I-systems. For cognitive systems, I-non-locality means that ideas which are separated in a p-adic space can be correlated. However, p-adic separation means only that there are no strong associations between ideas or groups of ideas. But this absence of associations does not imply that these ideas could not interact.

By our model each human society S has a wave function $\psi(t,x)$. This function gives the description of a quantum potential Q. The quantum potential can essentially change I-motions (i.e., evolutions of ideas) of individuals. Different societies are characterized by quantum potentials of different forms. This model provides an explanation of such collective phenomena as religion or political (or national) ideology.

The same considerations can be applied to animals and plants. The only difference is probably that here quantum I-potentials are not so strong. Thus we get the conclusion that there may exist the wave function $\psi_{\rm liv}(t,x)$ of all living organisms. The wave function $\psi(t,x)_{\rm liv}(t,x)$ can be represented in the

[14]We do not assume that $\psi(t,x)$ gives the complete description of τ. This description might be provided on the basis of hidden variables models which describe internal I-processes in τ (see, for example, [65], [66]). Of course, the $\psi(t,x)$ induces the definite trajectory in phase I-space $\mathbf{Z}_p^{2m}$. However, on the basis of the $\psi(t,x)$ we cannot describe internal I-processes in τ. In particular, an answer to the question Why does τ induce an I-field of the form $\psi(t,x)$? could not be given by the quantum formalism.

form:

$$\psi(t,x)_{\text{liv}}(t,x) = \sum_f \psi_f(t,x),$$

where $\psi_f(t,x)$ is a wave function of the living form f. An observable F (a living form) can be represented by a symmetric operator in a p-adic Hilbert space, $F\psi_f = f\psi_f$, where $f \in \mathbf{Z}_p$ is the cod of the living form f in the alphabet $\{0, 1, ..., p-1\}$. By equation (6.1) the evolution of the fixed form f_0 depends on evolutions of all living forms f.

The process of the evolution of living forms is not just a process based on Darwin's natural selection. This is a process of a quantum I-evolution in that the conscious field of all living forms plays the important role. This model might be used to explain some phenomena which could not be explained by Darwin's theory. For example, the beauty of colours of animals, insects and fishes could not be a consequence of the only process of the natural evolution. This is a consequence of the structure of the conscious field $\psi(t,x)_{\text{liv}}(t,x)$ [15]. By the same reasons we can explain some aspects of relations between robbers and victims. It seems that in nature there is a well organized system which gives to robers the right to eat victims. This system is nothing other than a result of the evolution due to (6.1).

Our formalism improves the ordinary pilot wave theory. One of delicate problems of this theory is the difference between ordinary fields and ψ-fields. There is no such a problem in the I-theory. All I-fields have no physical energy. Thus we need not discuss the energy balance for the field $\psi(t,x)$ (see [13], p.38). Of course, there is still the difference between classical and quantum rules for computing I-forces. As in the standard pilot wave theory, the increase of the (p-adic) amplitude of $\psi(t,x)$ does not imply the increase of the (p-adic) amplitude of the corresponding I-force. However, even in classical mechanics over p-adic numbers the increase of the amplitude of an I-force does not imply automatically an essential perturbation of a trajectory in phase I-space. For example, let $f = c \in \mathbf{Z}_p$. Then $p^{(c)}(t) = p_0 + c(t-t_0)$ and $q^{(c)}(t) = q_0 + p_0(t-t_0) + c/2(t-t_0)^2$. Thus $\Delta_p(t) = |p^{(c)}(t) - p^{(0)}(t)|_p = |c|_p |t-t_0|_p$ and $\Delta_q(t) = |q^{(c)}(t) - q^{(0)}(t)|_p = |c/2|_p |t-t_0|_p^2$.

[15] Of course, at the moment we cannot find such a function $\psi(t,x)_{\text{liv}}(t,x)$ which induces the real distribution of colours. In any case our model implies the existence of such a field. Therefore, the process of colours' evolution is a process of the simultaneous evolution of colours of numerous living forms. These colours do not serve only to the convenience of concrete forms (as it should be due to Darwin's theory), but they were produced by the correlated evolution of all living forms.

The quantities $\Delta_p(t)$ and $\Delta_q(t)$ can be very small for some instances of I-time. For example, if we choose a new time-scale $\tau_k = t_0 + kp^n$, where $k = 0, 1, ...$ and n is sufficiently large, then $\Delta_p(\tau_k)$ and $\Delta_q(\tau_k)$ will be practically zero. For example, if the I-state and motivation are measured at instances τ_k then we would not find a difference between the c-motion and free motion. Complex p-adic Hamiltonian systems have similar property. For example, let $f = cq, c \in \mathbf{Z}_p$, and let $p_0 = q_0 = 1$. Here $p^{(c)}(t) = e^{c(t-t_0)}$ and $\Delta_p(t) = |c|_p|t|_p$. Moreover, if we use the general formalism based on I-spaces over $\mathbf{Z}_m$, where m is not prime, then it may occur that $|ct|_p = 0$ for some instants of I-time.

Cognitive considerations imply a new viewpoint on the origin of the quantum I-field $\psi(t, x)$. According to BH (see especially [13], p. 38), the field $\psi(t, x)$ is an external field which guides the quantum particle. However, for a cognitive system τ (or a group of cognitive systems $\tau_1 ..., \tau_N$), it seems that this field is generated by τ (or by $\tau_1 ..., \tau_N$). Thus the interaction between an I-transformer and the field $\psi(t, x)$ is an interaction with self-induced I-field.

Moreover, the $\psi(t, x)$ can be considered as a new I-transformer, C. The C gets information from I-transformers $\tau_1, ..., \tau_N$ (via the classical field V) and C changes I-states of $\tau_1, ..., \tau_N$ (via equation (6.1)). The only distinguishing feature is that the I-state of C cannot be identified with a point in the I-space $\mathbf{Z}_p^m$ for a finite m. However, if we extend the I-formalism by using infinitely dimensional I-spaces over $\mathbf{Z}_p$, then C can be considered as an I-transformer. Thus each cognitive system (or a group of cognitive systems) induces a new I-transformer C (the consciousness) which evolves in infinite dimensional I-space.

In principle, C may induce a new field $\Phi(t, \psi(\cdot))$. This field determine the quantum potential for C. The field $\Phi(t, \psi(\cdot))$ can be again considered as an I-transformer C^1(which evolves in infinite dimensional I-space). It is the consciousness of (the consciousness C). If such a construction can be repeated many (or infinitely many?) times, then there arises a 'conscious tower', $C, C^1, ..., C^n, ...$. We might speculate that the motion of a cognitive system in I-space is determined by the hierarchic conscious system $C, C^1, ..., C^n, ...$. Of course, the effect of C^1 does not present in the linear Schrödinger equation. If we assume the hypothesis on the hierarchic conscious structure, then the linear Schrödinger equation has to be changed to nonlinear equation. Hence the cognitive considerations support de Broglie's ideas about nonlinear perturbations of Schrödinger's equation.

Remark 6.2. (Complex inner structure of quantum systems) One of the main consequences of BH considerations is that a quantum system (for

example, an electron) has a complex inner structure. Our consideration of quantum cognitive systems supports this idea. At the moment the existence of a complex inner structure for ordinary quantum systems is not strongly motivated. On the other hand, there is no doubt that such a structure exists in cognitive systems.

In our quantum formalism on I-spaces there is no difference between cognitive and non-cognitive quantum systems (in particular, each such a system has the I-field $\psi(t,x)$). Thus ordinary quantum particles might be considered as cognitive systems. From this point of view, for example, the two slit experiment demonstrates nothing other than cognitive behaviour of quantum particles.

7 Subjective probability as probability with respect to the statistical ensemble of human ideas

Let τ be an I-transformer representing the human brain. Here information space X_I is space of all ideas of τ. We propose the following ensemble interpretation of the subjective probability theory. The space of ideas X_I is the statistical ensemble which is used by τ for finding subjective probabilities. Let A be some event. Then A can be represented by τ as a subset of the statistical ensemble X_I. To find subjective probability, τ calculates the proportion:

$$\mathbf{P}^{\mathrm{sub}}(A) \equiv \mathbf{P}_{X_I}(A) = \frac{|A|}{|X_I|}\ .$$

Let τ use the coding system with m letters, $A_m = \{0, 1, ..., m-1\}$, and sequences $q = (\alpha_0, ..., \alpha_{N-1}), \alpha_j \in A_m$, of the length N for representing of ideas. If space of ideas X_I of τ contains all possible information strings of the length N, then $\mathbf{P}^{\mathrm{sub}}(\{a\}) = 1/m^N$ for each single idea $a \in X_I$.

On the other hand, in a mathematical model we can assume that ideas of τ are represented by information strings of the infinite length. If we suppose that space of ideas X_I of τ contains all infinite strings, then X_I can be identified with the set of m-adic integers $\mathbf{Z}_m$. To define conventional subjective probabilities, we use the uniform distribution μ on $X_I = \mathbf{Z}_m$. The μ is uniquely determined by its values on balls of $\mathbf{Z}_m$: $\mu(U_{1/m^k}(a)) = 1/m^k$ for any $a \in \mathbf{Z}_m$ and $k \geq 1$. In fact, this is the translation invariant measure

(the Haar measure) on the additive locally compact group $\mathbf{Z}_m$. Denote by $\mathcal{B}(\mathbf{Z}_m)$ the σ-algebra of Borel subsets of $\mathbf{Z}_m$. Thus (in the framework of real analysis) subjective probability theory can be described by the Kolmogorov probability space

$$\mathcal{P} = (\mathbf{Z}_m, \mathcal{B}(\mathbf{Z}_m), \mathbf{P}^{\rm sub}), \text{ where } \mathbf{P}^{\rm sub} = \mu.$$

To find subjective probability of an event $A \in \mathcal{B}(\mathbf{Z}_m)$ (a subset of space of ideas X_I of τ), τ performs the integration on A : $\mathbf{P}^{\rm sub}(A) = \int_A d\mu(x)$.

Some τ can have spaces of ideas X_I which are proper subsets of $\mathbf{Z}_m$. In such a case τ performs the integration $\mathbf{P}^{\rm sub}(A) = \int_{A\cap X_I} d\mu(x)$.

Remark 7.1. Of course, in some situations the brain can use nonuniform probability distributions on the statistical ensemble of ideas. Let ρ : $\mathbf{Z}_m \to \mathbf{R}_+$ be some probability density (namely, $\int_{\mathbf{Z}_m} \rho(x) d\mu(x) = 1$). Then $\mathbf{P}^{\rm sub}(A) = \int_{A\cap X_I} \rho(x) d\mu(x)$.

The great role of Bayes theorem (see Chapter 1) in subjective probability theory has a cognitive explanation. Consider a set of hypotheses $H_i, i = 1, ..., N$, (subsets of space of ideas X_I). Typically these sets of ideas have rather simple structure. As usual, we suppose that $H_i \cap H_j = \emptyset, i \neq j$, and

$$\cup_{j=1}^N H_j = X_I.$$

Let E be some event (a subset of space of ideas X_I) with rather complicated structure. But we suppose that implications $x \in H_i \to x \in E$ are easily verified for all $i = 1, ..., N$ (if an idea $x \in H_i$, then it easy to check the condition $x \in E$). Therefore conditional (subjective) probabilities $\mathbf{P}(E/H_i) = \mathbf{P}_{H_i}(E)$ can be easily found by τ as probabilities with respect to the statistical ensembles H_i. On the other hand, it is not so easy to find conditional probabilities $\mathbf{P}(H_i/E) = \mathbf{P}_E(H_i)$, because it is not so easy to verify the implication $x \in E \to x \in H_i$ (the statistical ensemble E has quite complicated structure and it is not easy to check even the condition $x \in E$). To find probabilities $\mathbf{P}(H_i/E)$, τ uses Bayes' theorem:

$$\mathbf{P}^{\rm Bayes}(H_i/E) = \frac{\mathbf{P}(E/H_i)\mathbf{P}(H_i)}{\sum_j \mathbf{P}(E/H_j)\mathbf{P}(H_j)} . \tag{7.1}$$

8 Bayes' theorem and Freud's psychoanalysis

The idea that Bayes' theorem (7.1) is used by a brain τ to calculate rather complicated conditional probabilities may be used as a probability explana-

tion of Freud's psychoanalysis. As in the previous section, we suppose that τ has a processor π_B which calculates probabilities $\mathbf{P}^{\text{Bayes}}(H_i/E)$ via (7.1). It may be that in some situations (by some psychological reasons) τ may check the results of functioning of π_B by calculating probabilities $\mathbf{P}(H_i/E)$ directly as probabilities with respect to the ensemble of ideas and by comparing two kinds of probabilities, $\mathbf{P}^{\text{Bayes}}(H_i/E)$ and $\mathbf{P}(H_i/E)$. Suppose that τ observes that $\mathbf{P}^{\text{Bayes}}(H_i/E)$ essentially differs from $\mathbf{P}(H_i/E)$. Such a situation is quite surprising for τ. The τ may start to calculate directly probabilities $\mathbf{P}(H_i/E)$ again and again for many different events E. It may be that the τ observes that for some class $\mathcal{E}$ of events E : $\mathbf{P}^{\text{Bayes}}(H_i/E) \neq \mathbf{P}(H_i/E)$. Such a situation may imply a psychological crisis of τ and even a mental decease. Moreover, it may even imply a physical decease. This has a simple probabilistic explanation. As a consequence of the negative experience with Bayes' probabilities $\mathbf{P}^{\text{Bayes}}(H_i/E)$ for $E \in \mathcal{E}$, the τ becomes to be afraid to use the processor π_B to calculate conditional probabilities with respect to events $E \in \mathcal{E}$. Instead of simple calculations (7.1), the τ must analyse the relation

$$x \in E \to x \in H_i \tag{8.1}$$

to find directly conditional probabilities $\mathbf{P}(H_i/E)$ as the ensemble probabilities

$$\mathbf{P}_E(H_i) = \frac{|E \cap H_i|}{|E|} .$$

If relation (8.1) is very complicated, then the calculation of $\mathbf{P}_E(H_i)$ takes a lot of time, energy and brain's memory. The τ could not obtain the decision for a reasonable time. The repeating of such a situation may imply that the τ becomes to be afraid to obtain any definite decision. This behaviour is nothing than *depression.* On the other hand, the τ may be involved in the practically infinite chain of considerations to analyse the structure of some concrete event $E_{\text{fix}} \in \mathcal{E}$. Such a behaviour is nothing than *manic behaviour.*

What is the origin of the difference between probabilities $\mathbf{P}^{\text{Bayes}}(H_i/E)$ (calculated via (7.1)) and $\mathbf{P}(H_i/E) = \mathbf{P}_E(H_i)$?

We think that Freud's theory on *conscious* and *unconscious* ideas gives the answer to this question. By Freud space of ideas X_I can be represented as

$$X_I = X_{I,c} \cup X_{I,u},$$

where $X_{I,c}$ and $X_{I,u}$ are spaces of conscious and unconscious ideas respectively. It is natural to suppose that τ uses hypotheses $H_1, ..., H_N$ consciously.

Thus $H_i \subset X_{I,c}, j = 1, ..., N$. On the other hand, it may be that $E \cap X_{I,u} \neq \emptyset$ for many $E \subset X_I$ and, moreover, it may be that the 'volume' $|E \cap X_{I,u}|$ is rather large. The conditional probabilities $\mathbf{P}(E/H_i)$ are based on the implication $x \in H_i \rightarrow x \in E$. Here τ uses only conscious ideas. Thus the calculation of $\mathbf{P}(E)$ via

$$\mathbf{P}(E) = \sum_{j=1}^{N} \mathbf{P}(E/H_j)\mathbf{P}(H_j) \tag{8.2}$$

is also based only on conscious ideas. In fact, (8.2) gives the ensemble probability $\mathbf{P}_{X_{I,c}}(E) = \mathbf{P}(E \cap X_{I,c})$ (in fact, $\cup_{j=1}^{N} H_j = X_{I,c}$). Hence

$$\mathbf{P}^{\text{Bayes}}(H_i/E) = \frac{\mathbf{P}(E \cap H_i)}{\mathbf{P}(E \cap X_{I,c})}.$$

Thus $\mathbf{P}^{\text{Bayes}}(H_i/E) = \mathbf{P}_{E \cap X_{I,c}}(H_i)$. On the other hand, we have

$$\mathbf{P}(H_i/E) = \mathbf{P}_E(H_i) = \frac{\mathbf{P}(E \cap H_i)}{\mathbf{P}(E \cap X_{I,c}) + \mathbf{P}(E \cap X_{I,u})}.$$

If $\mathbf{P}(E \cap X_{I,u})$ is rather large, then $\mathbf{P}(H_i/E) << \mathbf{P}^{\text{Bayes}}(H_i/E)$.

The aim of Freud's psychoanalysis is to find the unconscious component of the event E, $E_u = E \cap X_{I,u}$, and to move this component into space of conscious ideas $X_{I,c}$ of τ. The unconscious component E_u can contain a large amount of meaningless (from the conscious viewpoint) ideas and associations. By reading Freud's books it is impossible to understand why the conscious realization of such 'meaningless' information may really help to solve psychological problems of τ (and, in particular, to cure some mental deceases; for example, depression). Our probability model (based on probabilities with respect to ensembles of conscious and unconscious ideas) gives the explanation of Freud's method for the treatment of mental deceases: the conscious meaning of ideas $x \in E_u$ does not play any role, only the 'volume' of the set E_u in the space of ideas is important. After the mental treatment (via Freud's psychoanalysis) the space of conscious ideas $X_{I,c}$ is extended, $X_{I,c} \rightarrow X'_{I,c}$, in such a way that $E \subset X'_{I,c}$ (or at least $|E \cap X'_{I,u}|$ becomes small; here $X'_{I,u} = X_I \setminus X'_{I,c}$). Thus conditional probability $\mathbf{P}^{\text{Bayes}}(H_i/E)$ (calculated via (7.1)) becomes practically equal to conditional probability $\mathbf{P}(H_i/E) = \mathbf{P}_E(H_i)$. After some attempts of the comparative analysis of

these probabilities the τ understand that deviations of probabilities have been eliminated by Freud's psychoanalysis. The τ starts again to use Bayes formula (7.1) to calculate probabilities $\mathbf{P}(H_i/E)$. This implies the great economy of time, energy and memory resources.

Remark 8.1. Of course, our probabilistic model could not explain the whole mechanism of mental deviations induced by splitting of space of ideas X_I in a conscious and unconscious components. From the probabilistic viewpoint the 'volume' $|E_u|$ of the unconscious component E_u plays the crucial role. However, it may be that τ does not try whenever to check the result $\mathbf{P}^{\text{Bayes}}(H_i/E)$ (obtained via (7.1)) by the direct calculation of the ensemble probability $\mathbf{P}(H_i/E)$. Such a τ lives in the psychological world determined by Bayes' probabilities (7.1) and it would never have psychological problems. Our model could not explain why some individuals τ start to check (7.1) and analyse (8.1) and other individuals are not interested in such considerations. Our model cannot explain a mechanism of the exchange of ideas between conscious and unconscious components (see [68], [69] for some other models).

Chapter 6

Tests for randomness for p-adic probability theory

As it has been mentioned in Chapter 3, the first p-adic probability models [56], [60] were attempts to extend R. von Mises frequency probability theory to the p-adic case (see Chapter 4, section 2). As relative frequencies $\nu_N = \frac{n}{N} \in \mathbf{Q}$, we can study their behaviour not only in $\mathbf{R}$, but also in $\mathbf{Q}_p$. It is well know that von Mises' theory is based on two principles:
(1) the principle of the statistical stabilization of relative frequencies and (2) the principle of randomness. As we have seen, the first principle can be naturally generalized to the p-adic case and p-adic probabilities are defined as limits of relative frequencies with respect to p-adic topology. However, as in the ordinary real probability theory, there is the large problem with the principle of randomness. In the p-adic case the situation with stability of limits of relative frequencies with respect to place selections is even worse than in the real case, because the p-adic metric is very unstable: if $|n|_p = \varepsilon < 1$, then $|n+1|_p = 1$. In the p-adic case we have not even the possibility to restrict our considerations to a countable number of place selections (as we can do in the real case by Tornier theorem). To obtain the reasonable definition of p-adic randomness, we tried also to apply the theory of algorithmic complexity (see, for example, [18], [76], [77], [100], [79]). However, there was no large progress, see [60], [53], [54]. We present now a p-adic generalization of Martin-Löf's theory [83]–[85] based on tests for randomness[1]. Such a generalization looks as the most natural approach to p-adic randomness. Here

[1]This theory was developed by A. Khrennikov and S. Yamada [72].

we find natural tests for randomness for p-adic valued uniform probability distribution. Each test for randomness induces a series of limit theorems. On the other hand, individual limit theorems are not good candidates for test for randomness, because each theorem describes behavior of a subsequence $S_{n_k}(\omega)$ of the sequence $S_n(\omega) = \xi_1(\omega) + \cdots + \xi_n(\omega)$ of independent equally distributed random variables.

We proved that it is possible to enumerate effectively all p-adic test for randomness. However, in the opposite to Martin-Löf's theorem for real probabilities we proved that a universal p-adic test for randomness does not exist.

We shall use the standard terminology of the book of M. Li and P. Vitànyi [79]. The abbreviation r.e. is used for "recursive enumeration."

1 p-adic probability measures on the space of binary sequences

We set $X = \{0,1\}$, $X^n = \{x = (x_1, \ldots, x_n) : x_j \in X\}$, $X^* = \bigcup_n X^n$, $X^\infty = \{\omega = (\omega_1, \ldots, \omega_n, ...) : \omega_j \in X\}$. For $x \in X^n$, we set $l(x) = n$. For $x \in X^*$, $l(x) = n$, we define a cylinder U_x with basis x by $U_x = \{\omega \in X^\infty : \omega_1 = x_1, \ldots, \omega_n = x_n\}$. We denote by the symbol $\mathcal{F}_{\rm cyl}$ an algebra of subsets of X^∞ generated by all cylinders.

The map $j \colon X^\infty \to \mathbf{Z}_2$, $j(\omega) = \sum_{j=0}^\infty \omega_j 2^j$, gives one to one correspondence between X^∞ and $\mathbf{Z}_2$. Thus we can identify these sets. The algebra of cylindric sets $\mathcal{F}_{\rm cyl}$ coincides with the algebra $B(\mathbf{Z}_2)$ of all clopen subsets of $\mathbf{Z}_2$ (see Chapter 4).

A function $\mu \colon \mathcal{F}_{\rm cyl} \to \mathbf{Q}_p$ is a p-adic (valued) measure if the following properties holds true: (i) additivity: $\mu(A \cup B) = \mu(A) + \mu(B)$, $A \cap B = \emptyset$, $A, B \in \mathcal{F}_{\rm cyl}$; (ii) boundedness: $\|\mu\|_p = \sup\{|\mu(A)|_p : A \in \mathcal{F}_{\rm cyl}\} < \infty$. As it has been noticed in Chapter 4, the condition of continuity (iii) is redundant as $\mathcal{F}_{\rm cyl} = B(\mathbf{Z}_2)$.

A function $f \colon X^* \to \mathbf{Q}_p$ is said to be *recursive* iff there is a recursive function $g \colon X^* \times \mathbf{N} \to \mathbf{Q}$ such that $|f(x) - g(x,k)|_p < \frac{1}{k}$. A p-adic measure $\mu \colon \mathcal{F}_{\rm cyl} \to \mathbf{Q}_p$ is said to be recursive iff the function $f_p \colon X^* \to \mathbf{Q}_p$, $f_p(x) = \mu(U_x)$, is recursive.

The uniform p-adic measure μ_p ($p \neq 2$) on X^∞ is defined by

$$\mu_p(U_x) = \frac{1}{2^{l(x)}},\ x \in X^*. \tag{1.1}$$

If X^* is realized as $\mathbf{Z}_2$ and $\mathcal{F}_{\rm cyl}$ as $B(\mathbf{Z}_2)$, then μ_p is the p-adic valued Haar measure (translation invariant measure) on $\mathbf{Z}_2$.

As $\left|\frac{1}{2^{l(x)}}\right|_2 = 2^{l(x)}$, the additive set function μ_2 defined by (1.1) is not bounded. Therefore **we shall consider only the case** $p \neq 2$.

The simple considerations show that the function

$$N_{\mu_p}(x) = \inf\{\|U\|_\mu : x \in U \in B(\mathbf{Z_2})\} = 1$$

for all $x \in \mathbf{Z}_2$. This implies that $L_1(\mathbf{Z}_2, \mu_p) = C(\mathbf{Z}_2)$ (because all $B(\mathbf{Z}_2)$-step functions are continuous and each continuous function can be uniformly approximated by a sequence of $B(\mathbf{Z}_2)$-step functions). This implies that the algebra $(B(\mathbf{Z}_2))_{\mu_p} = \{A \subset \mathbf{Z}_2 : I_A \in L_1(\mathbf{Z}_2, \mu_p)\} = B(\mathbf{Z}_2)$. Thus the Haar measure μ_p cannot be extended from the algebra $B(\mathbf{Z}_2)$ to any larger algebra. In particular, the μ_p cannot be extended on the Borel σ-algebra generated by the algebra of clopen subsets $B(\mathbf{Z}_2)$.

Let a measure $\mu : \mathcal{F}_{\rm cyl} \to \mathbf{Q}_p$ be normalized: $\mu(X^\infty) = 1$. Then we can consider the p-adic probability space $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P})$, where $\Omega = X^\infty$, $\mathcal{F} = (\mathcal{F}_{\rm cyl})_\mu$(the set algebra which is obtained via the μ-extension of $\mathcal{F}_{\rm cyl}$), $\mathbf{P} = \bar{\mu}$ is a p-adic probability measure. As for the p-adic uniform measure μ_p (the $\mathbf{Q}_p$-valued Haar measure on $\mathbf{Z}_2$) the extension $(\mathcal{F}_{\rm cyl})_{\mu_p}$ coincides with $\mathcal{F}_{\rm cyl}$ and the extension $\bar{\mu}_p$ coincides with μ_p, the corresponding probability space is $\mathcal{P} = (\Omega, \mathcal{F}, \mathbf{P}_p)$, where $\Omega = X^\infty$, $\mathcal{F} = \mathcal{F}_{\rm cyl}$ and $\mathbf{P}_p = \mu_p$. The $\mathbf{P}_p$ is called a *uniform p-adic probability distribution.* We remark that values of $\mathbf{P}_p$ on cylinders coincide with values of the standard (real-valued) uniform probability distribution $\mathbf{P}_\infty$ on X^∞. As $\mathbf{Q} \subset \mathbf{R}$ and $\mathbf{Q} \subset \mathbf{Q}_p$, we can interpret rational numbers $\frac{1}{2^{l(x)}}$ both as real and as p-adic numbers.

In fact, we shall not use general recursive p-adic probabilities (see only definitions). We shall consider only the uniform p-adic probability distribution $\mathbf{P}_p$, $p \neq 2$ (which is, of course, recursive).

2 Some technical p-adic results

The results which are obtained in this section will be used to construct p-adic tests and prove limit theorems for p-adic probabilities.

For any $n, k \in \mathbf{N}$, (n, k) denotes the greatest common divisor of n and k; for any $n \in \mathbf{N}$, $M_p(n)$ denotes the mod p residue of $n : \ n = M_p(n)$ mod p. We set

$$\Theta_p(n) = \begin{cases} |n - M_p(n)|_p, & n \geq p. \\ 1, & 1 \leq n \leq p-1. \end{cases}$$

Lemma 2.1. *Let $n, k \in \mathbf{N}$, $k \leq n$ and let $M_p(n) \geq M_p(k)$. Then*

$$\left|\binom{n}{k}\right|_p = \frac{\Theta_p(n)}{\Theta_p(k)}.$$

Proof. Let $n = \alpha + ip^N$, $k = \beta + jp^l$, where $0 \leq \alpha, \beta \leq p-1$, $i, j, N, l \in \mathbf{N}$ and $(i,p) = (j,p) = 1$. We have:

$$\begin{aligned} \left|\binom{n}{k}\right|_p &= \left|(ip^N) \cdot \tfrac{(ip^N-p)}{p} \cdot \tfrac{(ip^N-2p)}{2p} \cdots \tfrac{(ip^N-jp^l+p)}{(jp^l-p)} \cdot \tfrac{1}{(jp^l)}\right|_p \\ &= \left|\tfrac{p^N}{p^l}\right|_p = p^{l-N}. \end{aligned} \tag{2.1}$$

To obtain (2.1), we have used that $n - k + 1 = ip^N - jp^l + (\alpha + 1 - \beta)$ and $0 < \alpha + 1 - \beta \leq p$; hence the last term in the nominator of $\binom{n}{k} = \frac{n \cdots (n-k+1)}{1 \cdots k}$, which is divisible by p is $(ip^N - jp^l + p)$. The cases in that $n = \alpha$ or $k = \beta$, $0 \leq \alpha$, $\beta \leq p - 1$ are considered in the same way. ■

Lemma 2.2. *Let $n, k \in \mathbf{N}$, $k \leq n$, and let $M_p(n) + 1 \leq M_p(k)$. Then*

$$\left|\binom{n}{k}\right|_p = \Theta_p(n).$$

Proof. Let $n = \alpha + ip^N$, $k = \beta + jp^l$, where $(i,p) = (j,p) = 1$, $0 \leq \alpha$, $\beta \leq p - 1$. We have

$$\begin{aligned} \left|\binom{n}{k}\right|_p &= \left|(ip^N) \cdot \tfrac{(ip^N-p)}{p} \cdot \tfrac{(ip^N-2p)}{2p} \cdots \tfrac{(ip^N-jp^l)}{(jp^l)}\right|_p \\ &= \left|p^N\right|_p = p^{-N}. \end{aligned} \tag{2.2}$$

To obtain (2.2), we have used that $n - k + 1 = (ip^N - jp^l) - (\beta - \alpha - 1)$ and $0 \leq \beta - \alpha - 1 < p$; hence the last term in the nominator of $\binom{n}{k} = \frac{n \cdots (n-k+1)}{1 \cdots k}$ which is divisible by p is $(ip^N - jp^l)$. The cases in that $n = \alpha$ or $k = \beta$, $0 \leq \alpha$, $\beta \leq p - 1$, are considered in the same way. ■

3 p-adic tests for randomness

We use the following notations. For each set $M \subset X^*$, we set $M^{(n)} = \{x \in M : l(x) = n\}$, $n = 1, 2, \ldots$. For each set $W \subset X^* \times \mathbf{N}$, we set $W_m = \{x \in X^* : (x,m) \in W\}$. Thus $W_m^{(n)} = \{x \in X^* : l(x) = n, (x,m) \in W\}$.

Everywhere in this chapter the **cardinality of a (finite) set A is denoted by the symbol $\sigma(A)$**. We do not use the standard symbol $|A|$, because we do not want to use expressions of the form $|\,|A|\,|_p$.

The following definition of a p-adic test for randomness is a natural generalization of Martin-Löf's definition of a test for randomness for ordinary real probabilities (in fact, in our particular case for the uniform distribution).

Definition 1. *Let $\mathbf{P}$ be a p-adic recursive probability. A recursively enumerable (r.e.) set $V \subset X^* \times \mathbf{N}$ is called a p-adic $\mathbf{P}$-test (p-adic test for randomness for the probability distribution $\mathbf{P}$) if it possesses the following two properties: for all $n, m \in \mathbf{N}$, we have:*

$$V_{m+1} \subset V_m,$$

$$\left| \sum_{x \in V_m^{(n)}} \mathbf{P}(U_x) \right|_p \le \frac{1}{p^m}. \tag{3.1}$$

The use of p-adic tests for randomness gives the possibility to formalize (in fact, to create) p-adic statistics. We are given the sample space X^* with an associated p-adic probability distribution $\mathbf{P}$. Given an element x of the sample space, we want to test hypothesis "x is a typical outcome". Practically speaking, the property of being typical is the property of belonging to reasonable majority. To ascertain whether a given element of the sample space belongs to a particular reasonable majority we use the notation of a test. As in the ordinary probability theory, a test is given by a prescription that, for every level of significance $\varepsilon = \frac{1}{p^m}$, tells us for what elements $x \in X^*$ the hypothesis "x belongs to majority M in X^*" should rejected where $\varepsilon = 1 - \mathbf{P}(M)$. The set V_m is a *critical region* on the *significance level* $\varepsilon = \frac{1}{p^m}$. If $x \in V_m$ then the hypothesis "x belongs to majority M" is rejected with the significance level ε. We say that x fails the test at the level of critical region V_m. Of course, there is a large difference between 'p-adic majority' and the ordinary 'real majority'. Populations which are very large from the point of view of ordinary real probability may be very small from the point of view of p-adic probability and vice versa.

We shall study **only the uniform p-adic probability distribution.** Everywhere below $\mathbf{P} = \mathbf{P}_p$, $p \neq 2$. Tests for randomness for this probability distribution we shall simply call p-adic test. Here condition (3.1) can be

reformulated in the following way:

$$\left|\sigma(V_m^{(n)})\right|_p \leq \frac{1}{p^m} \tag{3.2}$$

(as $\mathbf{P}(U_x) = \frac{1}{2^n}$ for $x \in V_m^{(n)}$ and $|2^n|_p = 1$ for $p \neq 2$, (3.1) has the form $\left|\sum_{x \in V_m^{(n)}} 1\right|_p \leq \frac{1}{p^m}$).

Proposition 3.1. *Let V be a p-adic test. Then, for each $(x, m) \in V$, we have*

$$l(x)(\log_p 2) > m \geq 1. \tag{3.3}$$

Proof. Set $n = l(x)$. As $x \in V_m$, we have $V_m^{(n)} \neq \emptyset$ and by (3.2) $\sigma(V_m^{(n)})$ is divisible by p^m. Thus $2^n = \sigma(X^n) \geq \sigma(V_m^{(n)}) \geq p^m$. This implies inequality (3.3). ■

Proposition 3.2. *Let V be a p-adic test. Then, for each $k \geq m$, $n \in \mathbf{N}$,*

$$\left|\sigma(V_m^{(n)} \setminus V_k^{(n)})\right|_p \leq \frac{1}{p^m}.$$

Proof. As $V_k^{(n)} \subset V_m^{(n)}$, we have:

$$\sigma(V_m^{(n)}) = \sigma(V_k^{(n)}) + \sigma(V_m^{(n)} \setminus V_k^{(n)}).$$

By the strong triangle inequality we get: $|\sigma(V_m^{(n)} \setminus V_k^{(n)})|_p \leq \max(|\sigma(V_m^{(n)})|_p,$ $|\sigma(V_k^{(n)})|_p) = 1/p^m$. ■

As usual, we denote the integer part of a real number x by $[x]$. Condition (3.3) can be rewritten in the form

$$[l(x)\log_p 2] \geq m.$$

The function $\lambda(n) = [n \log_p 2]$, $n \in \mathbf{N}$, will play the important role in our further considerations. For any p-adic test V and $n \in \mathbf{N}$, only sets $V_m^{(n)}$, $m = 1, \ldots, \lambda(n)$, can be nonempty.

We give now a few examples of p-adic tests for randomness. All these tests are related to behavior of sums:

$$S(x) = x_1 + \cdots + x_n, \quad x \in X^*, \quad n = l(x).$$

Example 3.1. We set

$$\begin{aligned} V_m = \{x \in X^* : \Theta_p(S(x)) \geq p^m \Theta_p(l(x)), S(x) \neq 0 \text{ and} \\ M_p(S(x)) \leq M_p(l(x))\}. \end{aligned} \tag{3.4}$$

To show that the set $V = \{(x, m) : x \in V_m\}$ is a p-adic test, we need only to show that (3.2) holds true. We have:

$$\sigma(V_m^{(n)}) = \sum_k \binom{n}{k},$$

where $0 \le k \le n$ and $M_p(k) \le M_p(n)$, $\frac{\Theta_p(n)}{\Theta_p(k)} \le \frac{1}{p^m}$. To obtain (3.2), it is sufficient to use the strong triangle inequality and Lemma 2.1.

Example 3.2. We set

$$\overline{V}_m = \left\{x \in X^* : \Theta_p(l(x)) \le \frac{1}{p^m} \text{ and } M_p(S(x)) \ge M_p(l(x)) + 1\right\}. \qquad (3.5)$$

By using Lemma 2.2 we obtain that (3.2) holds true for $\overline{V}_m$. Thus the set $\overline{V} = \{(x, m) : x \in \overline{V}_m\}$ is a a p-adic test.

Example 3.3. (Finite tests) Let $n \in \mathbf{N}$ be a fixed number. Let T be some subset of X^n, $\sigma(T) = p^{-\lambda(n)}$. We set $W_m^{(n)} = T$ for $m = 1, \ldots, \lambda(n)$ and $V_j^{(n)} = \emptyset$, $j > \lambda(n)$, and $V_j^{(n)} = \emptyset$, $k \neq n$, for all $j = 1, 2, \ldots$. Then $V = \{(x, m) : x \in V_m\}$, $V_m = \cup_{k=1}^{\infty} V_m^{(k)}$ is a finite p-adic test.

To illustrate the statistical meaning of tests (3.4) and (3.5), it is useful to consider some subsets of them corresponding to fixed values of $M_p(n)$ and $M_p(S(x))$.

We start with test (3.4). We set

$$\begin{array}{rcl} V_m(1,0) & = & \{x \in V_m : M_p(l(x)) = 1 \text{ and } M_p(S(x)) = 0\} \text{ and} \\ V(1,0) & = & \{(x, m) : x \in V_m(1,0)\}. \end{array} \qquad (3.6)$$

This test is connected with samples of the form

$$x = (x_1, \ldots, x_{1+jp^N}),\ j, N \in \mathbf{N},\ (j, p) = 1. \qquad (3.7)$$

Such a sample must be rejected with the level of significance $\varepsilon = \frac{1}{p^m}$ if $1 > |S(x)|_p \ge p^m |l(x) - 1|_p = p^{m-N}$. Thus the test $V(1,0)$ rejects all samples of the form $x = (x_1, \ldots, x_{1+jp^N})$, $(j, p) = 1$, in that the sum $S(x) = x_1 + \cdots + x_{1+jp^N}$ is not divisible by a sufficiently high degree of p (but divisible by p^1).

A sample x of form (3.7) with $S(x) = ip^k$, $(i, p) = 1, k \ge 1$, is rejected with the level of significance $\epsilon = 1/p^m$ if $k < N - m$.

For test (3.4) and $M_p(l(x)) = 1$, we can also fix $M_p(S(x)) = 1$ and obtain a new test:

$$\begin{array}{rcl} V_m(1,1) & = & \{x \in V_m : M_p(l(x)) = 1 \text{ and } M_p(S(x)) = 1\} \text{ and} \\ V(1,1) & = & \{(x,m) : x \in V_m(1,1)\}. \end{array}$$

A sample x of the form (3.7) must be rejected with the level of significance $\varepsilon = \frac{1}{p^m}$ if

$$1 > |S(x) - 1|_p \geq p^m |l(x) - 1|_p = p^{m-N}.$$

Thus the test $V(1,1)$ rejects all samples x of the form (3.7) for that $S(x)-1$ is not divisible by a sufficiently high degree of p (but divisible by p^1).

In the same way by fixing $M_p(n) = s = 0, \ldots, p-1$ we obtain tests $V_m(s,q)$, $q = 0, \ldots, s$. The $V(s,q)$ rejects some samples of the form

$$x = (x_1, \ldots, x_{s+jp^N}),\ j, N \in \mathbf{N},\ (j,p) = 1, \tag{3.8}$$

namely, samples for which $S(x) - q$ is not divisible by a sufficiently high degree of p (but divisible by p^1). A sample x of form (3.8) with $S(x) = q + ip^k$, $(i,p) = 1, k \geq 1$, is rejected with the level of significance $\epsilon = 1/p^m$ if $k < N - m$.

We study now test (3.5). The condition $M_p(S(x)) \geq M_p(l(x)) + 1 \geq 0$ implies that this test is used to reject (with some level of significance) some samples for that the sum $S(x)$ is not divisible by p (compare with (3.6)). We set

$$\begin{array}{rcl} \overline{V}_m(0,1) & = & \{x \in \overline{V}_m : M_p(l(x)) = 0 \text{ and } M_p(S(x)) = 1\} \text{ and} \\ \overline{V}(0,1) & = & \{(x,m) : x \in \overline{V}_m(0,1)\}. \end{array}$$

By this test we reject with the level of significance $\varepsilon = \frac{1}{p^m}$ all samples of the form $x = (x_1, \ldots, x_{jp^N})$, $(j,p) = 1$, for that $N < m$ and $M_p(S(x)) = 1$. We can compare the test $\overline{V}(0,1)$ with the test $V(0,0)$. The latter test is used to reject samples of the same form, but with $S(x)$ divisible by p: $S(x) = ip^k$, $(i,p) = 1, k \geq 1$. A sample is rejected with the level of significance $\varepsilon = \frac{1}{p^m}$ if $k < N - m$.

It is possible to introduce a p-adic test O which covers all cases of divisibility by p of $S(x)$. We start with the following simple fact:

Proposition 3.3. *Let Φ and Ψ be two p-adic tests such that $\Phi \cap \Psi = \emptyset$. Then the set $\Gamma = \Phi \cup \Psi$ is a p-adic test with critical regions $\Gamma_m = \Phi_m \cup \Psi_m$ on the significance level $\varepsilon = \frac{1}{p^m}$.*

Proof. We need only to prove that (3.2) holds true: We have $|\sigma(\Gamma_m^{(n)})|_p = |\sigma(\Phi_m^{(n)}) + \sigma(\Psi_m^{(n)})|_p \leq \max(|\sigma(\Phi_m^{(n)})|_p, |\sigma(\Psi_m^{(n)})|_p) \leq \frac{1}{p^m}$. ■

We now turn back to tests V and $\overline{V}$ defined in Examples 3.1, 3.2. It is evident that $V_m \cap \overline{V}_m = \emptyset$ for all m. Thus sets $\Sigma_m = V_m \cup \overline{V}_m$ give critical regions (with $\varepsilon = \frac{1}{p^m}$) of a p-adic test $\Sigma = \{(x, m) : x \in \Sigma_m\}$.

4 Some limit theorems

As in ordinary real probability theory tests V and $\overline{V}$ of Examples 3.1, 3.2 are related to some limit theorems for p-adic probability. Let $\mathcal{P} = (\Omega, \mathcal{F}_{cyl}, \mathbf{P})$ be the probability space based on the uniform p-adic distribution $\mathbf{P}$ on algebra $\mathcal{F}_{cyl}$ of cylindric subsets of $\Omega = X^\infty$, $p \neq 2$. For $\omega \in \Omega$, we set $S_n(\omega) = \omega_1 + \cdots + \omega_n$.

Theorem 4.1. *For each $l \in \mathbf{N}$ the probability*

$$\mathbf{P}\left(\left\{\omega \in \Omega : |S_n(\omega) - M_p(S_n(\omega))|_p = \frac{1}{p^l}, M_p(S_n(\omega)) \leq M_p(n)\right\}\right) \to 0$$

in $\mathbf{Q}_p$, when $|n - M_p(n)|_p \to 0$, $n \neq M_p(n)$.

Proof. By using considerations of Example 3.1 we obtain that

$$\mathbf{P}\left(\left\{\omega \in \Omega : |S_n(\omega) - M_p(S_n(\omega))|_p = \frac{1}{p^l}, M_p(S_n(\omega)) \leq M_p(n)\right\}\right)$$
$$\leq p^l |n - M_p(n)|_p.$$

■

In particular, we obtain the following limit theorems:

Corollary 4.1. *For each $l \in \mathbf{N}$, the probability*

$$\mathbf{P}\left(\left\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_{\frac{1}{p^l}}(0)\right\}\right) \to 0$$

in $\mathbf{Q}_p$, when $|n|_p \to 0$.

Corollary 4.2. (see [65]) *For each $l \in \mathbf{N}$, the probabilities*

$$\mathbf{P}\left(\left\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_{\frac{1}{p^l}}(0)\right\}\right) \text{ and } \mathbf{P}\left(\left\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_{\frac{1}{p^l}}(1)\right\}\right)$$

tend to zero in $\mathbf{Q}_p$, when $|n - 1|_p$ tends to zero.

Formally we can interpret Corollary 4.2 in the following way. The sum $S_n(\omega)$ can be considered as the sum $S_n(\omega) = \xi_1(\omega) + \cdots + \xi_n(\omega)$ of independent equally distributed random variables $\xi_j(\omega) = 0, 1$ with probabilities $1/2$. By Corollary 4.2 the probability distribution of random variable $S_{lim}(\omega) = \lim_{n\to 1} S_n(\omega)$ is concentrated at the points $a_0 = 0$ and a_1 of $\mathbf{Q}_p$. By symmetry reasons $\mathbf{P}_{S_{lim}}(\{a_0\}) = \mathbf{P}_{S_{lim}}(\{a_1\}) = 1/2$. Of course, this is just a formal statement, because Corollary 4.2 gives convergence only for spheres of $\mathbf{Q}_p$.

Theorem 4.2. *The probability*

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \geq M_p(n) + 1\}) \to 0$$

when $|n - M_p(n)|_p \to 0$.

As in the case of Theorem 4.1, we can, for example, put $M_p(n) = 0$ or $M_p(n) = 1$ and obtain the following consequences of Theorem 4.2:

Corollary 4.3. *The probability*

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \geq 1\}) \to 0$$

in $\mathbf{Q}_p$, *when* $|n|_p \to 0$.

Corollary 4.4. *The probability*

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \geq 2\}) \to 0$$

when $|n - 1|_p \to 0$.

We note that

$$\mathbf{P}(\{\omega \in \Omega : M_p(S_n(\omega)) \geq 1\}) = \mathbf{P}(\{\omega \in \Omega : S_n(\omega) \in \mathcal{S}_1(0)\}).$$

Thus by Corollaries 1 and 3 we obtain that

$$\mathbf{P}(\{\omega \in \Omega : S_n(\omega) \in \mathcal{U}_{\frac{1}{p^m}}(0)\}) \to 1,$$

$|n|_p \to 0$, for any $m \in \mathbf{N}$. Hence formally we obtain that the probability distribution $\mathbf{P}_{S_{lim}}$ of $S_{lim}(\omega) = \lim_{n\to 0} S_n(\omega)$ is concentrated at the point $a_0 = 0 \in \mathbf{Q}_p$, $\mathbf{P}_{S_{lim}}(\{0\}) = 1$.

It seems that in the p-adic case it is more natural to use tests for randomness than limit theorems. In the opposite to ordinary real probability theory in the p-adic case we have no general limit theorems for $n \to \infty$ (in the sense of the order on $\mathbf{N}$). All limit theorems give the convergence of probabilities for some sequences $n_k \to \infty$, $k \to \infty$. For example, $|n_k|_p \to 0$, $n_k \neq 0$, implies that $n_k = jp^N$, $(j, p) = 1$, $N \to \infty$, and $|n_k - 1|_p \to 0$, $n_k \neq 1$, implies that $n_k = 1 + jp^N$, $(j, p) = 1$, $N \to \infty$, and so on.

5 Recursive enumeration of the set of p-adic tests

Here we shall prove that the set of all p-adic tests is recursively enumerable. The general scheme of the proof is the same as in the case of real probabilities. However, the main part of the proof (an algorithm for constructing a p-adic test on the basis of a partial recursive function) strongly differs from the standard one (see [79]).

We start with the following well known Lemma (see, for example, [79]).

Lemma 5.1. *There exists a partial recursive function $f\colon \mathbf{N}\times\mathbf{N}\to X^*\times\mathbf{N}$ with the following properties:*

(a1) *for all $i,j\in\mathbf{N}$ such that $f(i,j)\neq\infty$, we have $f(i,k)\neq\infty$, for all $k\leq j$;*

(a2) *a set $A\subset X^*\times\mathbf{N}$ is r.e. iff $A=\{f(i,j):j=1,2,\dots\}\setminus\{\infty\}$, for some $i\geq 1$.*

Theorem 5.1. *The set of all p-adic tests is r.e. .*

Proof. Through the proof we shall use the fixed partial recursive function $\varphi=\varphi_i=f(i,\cdot)$ given by Lemma 1. We set $A_\varphi=\varphi(\mathbf{N})$. As in the standard case, we shall construct for each φ some total recursive function $g\colon \mathbf{N}\to X^*\times\mathbf{N}$ such that $T=A_g=g(\mathbf{N})$ is a p-adic test and if φ is a p-adic test by itself, then $T=A_\varphi$. We construct T step by step using an algorithm which produces a p-adic test at each step. In the following algorithm we shall use sets $\mathcal{D}_m^{(n)}$ which give approximations for sets $T_m^{(n)}$ in the process of building of T (as usual $T_m=\{x\in X^*:(x,m)\in T\}$ and $T_m^{(n)}=\{x\in T_m: l(x)=n\}$). We shall also use sets $R_m^{(n)}$ which are registers for collecting elements of $(A_\varphi)_m^{(n)}$. The main difference with the standard algorithm is due to the fact that we cannot increase sets $\mathcal{D}_m^{(n)}$ at each step when φ produces a value $\varphi(j)\in(A_\varphi)_m^{(n)}=\{x:\phi(j)=(x,m)$ for some j and $l(x)=n\}$ (because the p-adic metric is changed discontinuously: $|x|_p\leq\frac{1}{p^m}\Rightarrow|x+1|_p=1$, $m\geq 1$). We collect (in $R_m^{(n)}$) elements of $(A_\varphi)_m^{(n)}$ until $\sigma(R_m^{(n)})$ becomes divisible by p^m. After this we set $\mathcal{D}_m^{(n)}=R_m^{(n)}$.

To be sure that the result of our construction will be a r.e. set, we construct parallel a function $g\colon\mathbf{N}\to X^*\times\mathbf{N}$ such that $T=g(\mathbf{N})$ and g is a total recursive function if T is an infinite set.

Algorithm

1 Put $T = \emptyset$, $\mathcal{D}_m^{(n)} = R_m^{(n)} = \emptyset$; put $j = 0$, $i = 0$, $t_m^{(n)} = 0$.

% j is the argument of φ, i is the argument of g; $t_m^{(n)} = \sigma(R_m^{(n)})$.

2 Put $j = j + 1$

3 If $\varphi(j) = \infty$ continual indefinitely.

4 Find $\varphi(j) = (x, m)$ and $n = l(x)$.

5 If $m > [n \log_p 2]$, then $T = \emptyset$ and stop.

6 Put $R_m^{(n)} = R_m^{(n)} \cup \{x\}$ and $t_m^{(n)} = t_m^{(n)} + 1$.

7 If $|t_m^{(n)}|_p > \frac{1}{p^m}$, then go to step 2.

8 If $m \geq 2$ and $\mathcal{D}_{m-1}^{(n)} \not\supset R_m^{(n)}$, then go to step 2.

9 Put $\mathcal{D}_m^{(n)} = R_m^{(n)}$.

% We must make step 8 before step 9 to get $T_{m-1} \supset T_m$.

10 **(a)** Enumerate elements if $\mathcal{D}_m^{(n)} = \{z_1, \ldots, z_{t_m^{(n)}}\}$;

(b) for $l = 1, \ldots, t_m^{(n)}$, put $g(i + l) = (z_l, m)$;

(c) put $i = i + t_m^{(n)}$.

% The previous step is not related to the construction if T; here we construct the function g which gives recursive enumeration for T.

11 Put $s = m$.

12 Put $s = s + 1$.

13 If $s > [n \log_p 2]$, go to 18.

14 If $|t_s^{(n)}|_p > \frac{1}{p^s}$, go to 18.

15 If $\mathcal{D}_{s-1}^{(n)} \not\supset R_s^{(n)}$, go to 18.

16 Put $\mathcal{D}_s^{(n)} = R_s^{(n)}$.

% We explain the meaning of steps 11 - 16. By step 9 the set $\mathcal{D}_m^{(n)}$ has been increased. Thus condition 8 must be reconsidered for sets $\mathcal{D}_s^{(n)}$ with $s > m$. It can be that occasionally some of sets $R_s^{(n)}$ has the number of elements which is divisible by p^s. If they pass step 15, then we increase sets $\mathcal{D}_j^{(n)}$ by 16.

17 Repeat step 10 for $m = s$.

18 Put $T = T \cup_{m \leq s \leq [n \log_p 2]} \mathcal{D}_s^{(n)} \times \{s\}$ and go to step 2.

We prove now that the set T which is constructed by the algorithm is a p-adic test.

(A1) We use the parameter j to denote the step (determined by 2) of the algorithm. We have $T_m^{(n)} = \cup_{j=1}^{\infty} \mathcal{D}_m^{(n)}(j)$. As $\mathcal{D}_m^{(n)}(j+1) \supset \mathcal{D}_m^{(n)}(j)$ and $\sigma(\mathcal{D}_m^{(n)}(j))$ is divisible by p^m, we get that $\sigma(T_m^{(n)})$ is divisible by p^m. Thus $|\sigma(T_m^{(n)})|_p \leq \frac{1}{p^m}$.

(A2) By step 8 and 15 we get that $\mathcal{D}_m^{(n)} \supset \mathcal{D}_{m+1}^{(n)}$, $n, m \in \mathbf{N}$. Thus $T_m^{(n)} \supset T_{m+1}^{(n)}$, $n, m \in \mathbf{N}$.

(A3) If steps 10 and 16 are passed an infinite number of times, then g is the total recursive function and, hence, $T = A_\varphi$ is r.e.. If the steps 10 and 16 are passed only a finite number of times, then the set T is finite and, hence, r.e.

We prove now that if $V = A_\varphi$ is a p-adic test, then $T = A_\varphi$.

It is evident that $T \subset A_\varphi$. We have only to prove that $V \subset T$. It is sufficient to prove that, for each n, $V_m^{(n)} \times \{m\} \subset T_m^{(n)} \times \{m\}$ for all $m \leq [n \log_p 2]$.

For each n, the set $V^{(n)} = \{(x, m) \in V : l(x) = n\}$ is finite (since $m \leq [n \log_p 2]$). Thus φ produces all elements of $V^{(n)}$ after a finite numbers of steps $J = J(n, \varphi)$.[2] Let $\varphi(J) = (x_J, m_J)$ (here $l(x_J) = n$ and $m_J \leq [n \log_p 2]$). We have: $\mathcal{D}_1^{(n)} \supset \mathcal{D}_2^{(n)} \supset \cdots \supset \mathcal{D}_M^{(n)}$ and $|\sigma(\mathcal{D}_s^{(n)})|_p \leq \frac{1}{p^s}$ for $s = 1, \ldots, M = [n \log_p 2]$. We also have: $R_s^{(n)} = V_s^{(n)}$ (because $V^{(n)} \subset \varphi(\{1, 2, \ldots, J\})$ and

[2] Of course, some points $(x, m) \in V^{(n)}$ can appear again on some steps $J' > J$.

$\varphi(\{1,2,\ldots,J\})_s^{(n)} = R_s^{(n)}$). Thus, for all s, $|\sigma(R_s^{(n)})|_p \leq \frac{1}{p^s}$. In particular, this holds for $s = m_J$. Hence, for $m = m_J$, step 7 is passed.

We prove that $\mathcal{D}_s^{(n)} = R_s^{(n)} = V_s^{(n)}$ for all $s = 1, \ldots, m_J - 1$. As $|\sigma(R_1^{(n)})|_p \leq \frac{1}{p}$, $R_1^{(n)}$ has passed step 7. But step 8 is trivial for $m = 1$. Thus by step 9 we get $\mathcal{D}_1^{(n)} = R_1^{(n)} = V_1^{(n)}$. For $s = 2$, we have $|\sigma(R_2^{(n)})|_p \leq \frac{1}{p^2}$ and step 7 is passed. As $\mathcal{D}_1^{(n)} = V_1^{(n)}$ and $R_2^{(n)} = V_2^{(n)}$, we have $\mathcal{D}_1^{(n)} \supset R_2^{(n)}$ and step 8 is passed. By step 9 we get $\mathcal{D}_2^{(n)} = R_2^{(n)} = V_2^{(n)}$. We can repeat such considerations until s takes value $m_J - 1$.

As $\mathcal{D}_{m_J-1}^{(n)} = V_{m_J-1}^{(n)} \supset V_{m_J}^{(n)} = R_{m_J}^{(n)}$, step 8 is passed for $m = m_J$ and we get $\mathcal{D}_{m_J}^{(n)} = R_{m_J}^{(n)} = V_{m_J}^{(n)}$. Thus we arrive to step 11 with $m = m_J$. For all $m_J < s \leq M = [n \log_p 2]$, step 14 is passed automatically. For $s = m_J + 1$ we have $\mathcal{D}_{m_J}^{(n)} = V_{m_J}^{(n)} \supset V_{m_J+1}^{(n)} = R_{m_J+1}^{(n)}$. Hence step 15 is passed and we put $\mathcal{D}_{m_J+1}^{(n)} = R_{m_J+1}^{(n)} = V_{m_J+1}^{(n)}$. Repeating these considerations, we prove that $\mathcal{D}_s^{(n)} = R_s^{(n)} = V_s^{(n)}$ for all $s = m_J, \ldots, M$. Hence $V^{(n)} = T^{(n)}$. ∎

6 No p-adic universal test

A natural generalization of the definition of a universal test for randomness is the following one:

Definition 6.1. *A p-adic test U is said to be universal if for every p-adic test V we can effectively find $c \in \mathbf{N}$ (depending upon U and V) such that $V_{m+c} \subset U_m$ for all m.*

It is well known that in the ordinary real probability theory there exists a universal test for randomness (which is, of course, not unique). We shall show that in p-adic probability theory there is no universal recursive tests. We start with some technical considerations. We have to study more carefully properties of the function $\lambda(n) = [n \log_p 2]$. As $p > 2$, we have $\log_p 2 < 1$. We set $L_k = \left[\frac{k}{\log_p 2}\right]$. If $0 < n \leq L_1$, then $n \log_p 2 < 1$ and $\lambda(n) = 0$; in the same way we have: if $L_{k-1} < n \leq L_k$, then $\lambda(n) = k - 1$, $k \geq 2$. We set $n_k = L_k + 1$.

Lemma 6.1. *The inequality*

$$p^{\lambda(n_k)} > 2^{n_k - 1} \tag{6.1}$$

holds true for all $k = 1, 2, \ldots$.

Proof. We have $\lambda(n_k) = k$ and $\lambda(n_k - 1) = k - 1$. By definition $\lambda(n) = \max\{l : p^l < 2^n\}$. Hence, for all n, $p^{\lambda(n)+1} > 2^n$. In particular, $p^{\lambda(n_k-1)+1} = p^k > 2^{n_k-1}$. Hence $p^k = p^{\lambda(n_k)} > 2^{n_k-1}$. ∎

We construct now two p-adic tests W and $\widetilde{W}$ by using the following procedure.

For $k = 1, 2, \ldots$ and $j = 1, \ldots, \lambda(n_k)$, we set

$$W_j^{(n_k)} = W_{\lambda(n_k)}^{(n_k)} = \{x_1, \ldots, x_{p^{\lambda(n_k)}}\}$$

and

$$\widetilde{W}_j^{(n_k)} = \widetilde{W}_{\lambda(n_k)}^{n_k} = \{x_{2^{n_k}-p^{\lambda(n_k)}+1}, \ldots, x_{2^{n_k}}\}$$

and $W_l^{(n_k)} = \widetilde{W}_l^{(n_k)} = \emptyset$ for $n \neq n_k$ and $l = 1, 2, \ldots$. Here we have used the lexicographic enumeration of elements of X^{n_k}, $k = 1, 2, \ldots$: $x_1, x_2, \ldots, x_{2^{n_k}}$. Since $\sigma(W_j^{(n_k)}) = \sigma(\widetilde{W}_j^{(n_k)}) = p^{\lambda(n_k}$, $j = 1, 2 \ldots, \lambda(n)$, by (6.1) we obtain $W_j^{(n_k)} \cap \widetilde{W}_j^{(n_k)} \neq \emptyset$ and hence

$$X^{n_k} = W_j^{(n_k)} \cup \widetilde{W}_j^{(n_k)}, \quad j = 1, \ldots, \lambda(n_k).$$

Theorem 6.1. *A universal p-adic test does not exist.*

Proof. Let us suppose that there exists a universal p-adic test U. Thus we can effectively find $c_1, c_2 \in \mathbf{N}$ such that $W_{m+c_1} \subset U_m$ and $\widetilde{W}_{m+c_2} \subset U_m$, where W and $\widetilde{W}$ are p-adic tests constructed before this theorem. Let k be so large that $\lambda(n_k) - c_1 \geq 1$ and $\lambda(n_k) - c_2 \geq 1$. Thus $W_{1+c_1}^{(n_k)} = W_{\lambda(n_k)}^{(n_k)}$, $\widetilde{W}_{1+c_2}^{(n_k)} = \widetilde{W}_{\lambda(n_k)}^{(n_k)}$. Hence $U_1^{(n_k)} \supset W_{1+c_1}^{(n_k)} \cup \widetilde{W}_{1+c_2}^{(n_k)} = X^{n_k}$. This implies that $|\sigma(U_1^{(n_k)})|_p = |\sigma(X^{n_k})|_p = 1$. This contradicts to (3.2). ∎

7 Randomness of infinite sequences

Let $V \subset X^* \times \mathbf{N}$ be a p-adic $\mathbf{P}$-test, where $\mathbf{P}$ is an arbitrary p-adic recursive probability. We set

$$O_m = \cup\{\mathcal{U}_y : y \in V_m\} \subset X^\infty \text{ and } O = \{(\omega, m) : \omega \in O_m\} \subset X^\infty \times \mathbf{N}. \quad (7.1)$$

If the set V_m is infinite, then in general O_m does not belong to the set algebra $\mathcal{F}_{cyl}$. Therefore probability $\mathbf{P}(O_m)$ may be not defined.

Thus we could not generalize the standard condition for real probabilities (namely $\mathbf{P}(O_m) \leq \frac{1}{p^m}$) to define a p-adic sequential test. It seems that the

only possibility to define a p-adic sequential test is to use all tests O obtained via (7.1) from p-adic $\mathbf{P}$-tests $V \subset X^* \times \mathbf{N}$.

Definition 7.1. *Let $\mathbf{P}\colon \mathcal{F}_{cyl} \to \mathbf{Q}_p$ be a recursive probability and let $V \subset X^* \times \mathbf{N}$ be a p-adic $\mathbf{P}$-test. The set O defined on the basis of V via (7.1) is said to be a p-adic sequential $\mathbf{P}$-test.*

Definition 7.2. *Let O be a p-adic sequential test. A sequence ω is said to be $\mathbf{P}$-random with respect to the test O if*

$$\omega \notin O_\infty = \bigcap_{m=1}^{\infty} O_m.$$

In general, the set $O_\infty \notin \mathcal{F}_{cyl}$ and $\mathbf{P}$ cannot be extended on the σ-algebra $\mathcal{B}(X^\infty)$ containing O_∞. Therefore in general $\mathbf{P}(O_\infty)$ is not defined.

A sequence $\omega \in O_\infty$ is considered as non-random with respect to $\mathbf{P}$.

As usual, we restrict our considerations to the case of the uniform p-adic distribution $\mathbf{P} = \mathbf{P}_p$, $p \neq 2$.

Example 7.1. Let O be the p-adic sequential test based on the p-adic test V of Example 3.1. We consider a few examples of sequences $\omega \in X^\infty$ which are random or non-random with respect to O:

(1) Let only a finite number $k \geq 1$ of coordinates of $\omega = (\omega_j)$ be equal to 1: $\omega_{j_1} = \cdots = \omega_{j_k} = 1$. We show that ω is not non-random with respect to O. We have to show that for each m there exist n such that $\omega_{1:n} \in V_m^{(n)}$ where $\omega_{1:n} = (\omega_1, \ldots, \omega_n)$. Let $k = \beta = 1, \ldots, p-1$. We set $n = \beta + p^N$, where $N \geq m$, $N > \log_p(j_k - \beta)$. Then $\Theta_p(n)/\Theta_p(k) = p^{-N} \leq p^{-m}$ and $\omega_{1:n} \in V_m^{(n)}$. Let $k = \beta + jp^l$, $j, l \in \mathbf{N}$, $(j,p) = 1$. We set $n = \beta + p^N$, where $N \geq m + l$, $N > \log_p(j_k - \beta)$. Then $\Theta_p(n)/\Theta_p(k) = p^{-N+l} \leq p^{-m}$ and $\omega_{1:n} \in V_m^{(n)}$.

(2) The sequence $\omega = (0, \ldots, 0, \ldots)$ is random with respect to O, because $\omega \notin O_1$ ($\omega_{1:n} \notin V_1^{(n)}$ for all n); the sequence $\omega = (1, \ldots, 1, \ldots)$ is random with respect to O, because $k = S(\omega_{1:n}) = n$ and $\Theta_p(n)/\Theta_p(k) = 1$ and hence $\omega_{1:n} \notin V_1^{(n)}$ for all n.

(3) Here we present an example of a random sequence $\omega \in X^\infty$ with respect to O which contains the infinite number both of zeros and ones. Any $\omega \in X^\infty$ can be represented as a sequence of blocks $\omega = b_1 b_2 \ldots b_m \ldots$, where $l(b_j) = p^{2j}$. Let $S(b_1) = p$ and $S(b_j) = p^j - p^{j-1}$, $j > 1$. Set $x = b_1 \ldots b_m, m \geq 1$. Then $x \in V_m^{(p^{2m})}$. Here $l(x) = p^{2m}$ and $S(x) = p + (p^2 - p) + \cdots + (p^m - p^{m-1}) = p^m$, thus: $\Theta_p(l(x))/\Theta_p(S(x)) = p^{-m}$.

Example 7.2. Let $\overline{O}$ be a p-adic sequential test based on the p-adic test $\bar{V}$ of Example 3.2.

(1) We consider the same sequence ω as in (1) of Example 7.1.

(a) Let $k = \beta$ or $k = \beta + jp^l$, $(j, p) \neq 1$ and $\beta = 1, \ldots, p-1$. We show that such an $\omega \in X^\infty$ is non-random with respect to $\overline{O}$. Let $n = (\beta - 1) + p^N$, where $N \geq m$ and $N > \log_p(j_k - \beta + 1)$. Then $S(\omega_{1:n}) = k$ and hence $M_p(S(\omega_{1:n})) = \beta \geq M_p(n) + 1$ and $\Theta_p(n) = p^{-N} \leq p^{-m}$. Thus $\omega_{1:n} \in \overline{V}_m^{(n)}$ and $\omega \in \overline{O}_m$.

(b) Let $k = jp^l$, $(j, p) \neq 1$. We show that if $p^m \geq j_k$, then $\omega_{1:n} \notin V_m^{(n)}$, $n \geq 1$ (the condition $\Theta_p(l(x)) \leq p^{-m}$ implies that $l(x) \geq p^m$; but, for $\omega_{1:n}$ with $n \geq p^m$, we have $S(\omega_{1:n}) = k$ and, as $M_p(S(\omega_{1:n})) = 0$, there is no n such that $M_p(S(\omega_{1:n})) \geq M_p(n) + 1$). Thus in the opposite to the test O any sequence $\omega \in X^\infty$ in that only a finite number $k = p^t$, $t = 1, 2, \ldots$, of coordinates are equal to 1 is considered as random with respect to $\overline{O}$.

(2) The sequence $\omega = (0, \ldots, 0, , \ldots)$ is random with respect to $\overline{O}$ (because, for all $x = (0, \ldots, 0)$, $0 = M_p(S(x)) < M_p(l(x)) + 1$; the sequence $\omega = (1, \ldots, 1, \ldots)$ is also random with respect to $\overline{O}$ (because, for all $x = (1, \ldots, 1)$, $M_p(S(x)) = M_p(l(x))$).

(3) We consider the same sequence as in (3) of Example 7.1. We show that some of such sequences are random with respect to $\overline{O}$ and some are non-random. Let $\omega = b_1 b_2 \ldots b_m \ldots$ and in each block b_j the first $p^j - p^{j-1}$ elements are equal 1 and $\omega_1 = \cdots = \omega_p = 1$ in b_1 (other elements in each block are equal to 0). If, for $\omega_{1:n}$, $\Theta_p(n) \leq \frac{1}{p^m}$, then $M_p(n) = 0$ and, hence, $M_p(S(\omega_{1:n})) = 0$. Thus $\omega_{1:n} \notin \overline{V}_m^{(n)}$. Let $\omega = b_1 b_2 \ldots b_m \ldots$ and the distribution of ones in blocks have the following structure. For $b_1 = (x_1, \ldots, x_{p^2})$, $x_1 = \cdots = x_{p-1} = 1$, $x_p = 0$, $x_{p+1} = 1$; for $b_j = (x_1, \ldots, x_{p^{2j}})$, $x_1 = \cdots = x_{p^j - p^{j-1} - 1} = 1$, $x_{p^j - p^{j-1}} = 0$, $x_{p^j + p^{j-1} + 1} = 1$. Then $\omega_{1:p} \in V_1^{(p)}$ (since $M_p(S(\omega_{1:p})) = p - 1 > 1 + M_p(p) = 1$); $\omega_{1:n} \in V_{p-j+1}^{(n)}$ for $n = p^{2j} + p^j - p^{j-1}$ (since $S(\omega_{1:n}) = p^j - p^{j-1} - 1$ implies $M_p(S(\omega_{1:n})) = 1$).

As consequence of Theorem 5.1 we obtain the following theorem:

Theorem. 7.1. *The set of all p-adic sequential tests is r.e..*

A p-adic sequential test $\mathcal{D}$ is said to be universal if, for every p-adic sequential test O, we can effectively find $c \in \mathbf{N}$ (depending upon $\mathcal{D}$ and O) such that $O_{m+c} \subset \mathcal{D}_m$ for all m.

At first sight, it seems to be natural to consider the set

$$O_{\infty}^{max} = \bigcup_{i=1}^{\infty} O_{(i),\infty}, \tag{7.2}$$

where $O_{(i)}$, $i = 1, 2, \ldots$, is a recursive enumeration of p-adic sequential tests, as the maximal set of p-adic non-random sequences (with respect to the p-adic uniform distribution) and call a sequence $\omega \in X^{\infty} \setminus O_{\infty}^{max}$ a p-adic random sequence. However, Theorem 6.1 (nonexistence of universal p-adic test) is the sign that such a procedure could not be successful.

Proposition 7.1. *The set O_{∞}^{max} defined as (7.2) is equal to X^{∞}.*

Proof. Let W and $\widetilde{W}$ be p-adic tests defined in section 6 and let O and $\widetilde{O}$ be the corresponding p-adic statistical tests. We have that $W_j^{(n)} \cup \widetilde{W}_j^{(n)} = X^n$, $j = 1, \ldots, \lambda(n)$, for all n. As $\lambda(n) \to \infty$, $n \to \infty$, then $\forall l, m \in \mathbf{N}\ \exists N = N(l,m)$: $\lambda(N) \geq l, m$. As we also have that $W_j^{(n)} = W_{\lambda(n)}^{(n)}$ and $\widetilde{W}_j^{(n)} = \widetilde{W}_{\lambda(n)}^{(n)}$, $j = 1, \ldots, \lambda(n)$, then, for $N = N(l,m)$, we obtain $W_l^{(N)} \cup \widetilde{W}_m^{(N)} = X^N$. We also have:

$$O_{\infty} \cup \widetilde{O}_{\infty} = (\cap_{l=1}^{\infty} O_l) \cup (\cap_{m=1}^{\infty} \widetilde{O}_m) = \cap_{l=1}^{\infty} \cap_{m=1}^{\infty} (O_l \cup \widetilde{O}_m).$$

Finally we show that $O_l \cup \widetilde{O}_m = X^{\infty}$ for every l and m:

$$O_l \cup \widetilde{O}_m \supset \cup \{\mathcal{U}_x : x \in W_m^{(N)} \cup \widetilde{W}_l^{(N)}\} = X^{\infty},$$

where $N = N(l,m)$. ∎

Thus in the opposite the real case the existence of the recursive enumeration of the set of all p-adic sequential tests does not imply the possibility of the fruitful development of the theory of randomness based on the maximal constructive set of non-random sequences. In some sense the situation here is similar to the ordinary (real) nonconstructive probability theory where any $B \subset X^{\infty}$, $\mathbf{P}(B) = 1$, may be considered as a 'law of randomness' (thus the maximal set of non-radnom sequences coincides with X^{∞}).

Definition 7.2. *A p-adic sequential test $\mathcal{O}$ is said to be a universal if for every p-adic sequential test O we can effectively find $c \in \mathbf{N}$ such that $O_{m+c} \subset \mathcal{O}_m$ for all m.*

Lemma 7.1. *Let n_j, $j = 1, 2, \ldots$, be numbers associated with the function λ. Then, for $m_j = \lambda(n_j)$,*

$$N_j = (2^{n_j} - p^{m_j}) 2^{n_{j+1} - n_j} < p^{m_j + 1}, \quad j = 1, 2, \ldots .$$

Proof. By (6.1) we obtain: $p^{m_j}2^{-n_j} > 1/2$. Thus $(1 - p^{m_j}2^{-n_j}) < 1/2$ and hence

$$N_j = (1 - p^{m_j}2^{-n_j})2^{n_{j+1}} < \frac{2^{n_{j+1}}}{2}.$$

But by (6.1) we also have $p^{m_{j+1}} > \frac{2^{n_{j+1}}}{2}$. ∎

Proposition 7.2. *The trivial p-adic sequential test $\mathcal{O}$ with $\mathcal{O}_m = X^\infty$, $m \geq 1$, is the (unique) universal p-adic sequential test.*

Proof. We prove that the $\mathcal{O}$ with $\mathcal{O}_m = X^\infty$ for all $m \geq 1$ is a p-adic sequential test: there exists a p-adic test $V \subset X^* \times \mathbf{N}$ such that V induces $\mathcal{O}$. Let n_j be natural numbers associated with λ and let $m_j = \lambda(n_j)$.

We represent $X^{n_j} = A_{m_j}^{(n_j)} \cup B_{m_j}^{(n_j)}$, $A_{m_j}^{(n_j)} \cap B_{m_j}^{(n_j)} = \emptyset$ and $\sigma(A_{m_j}^{(n_j)}) = p^{m_j}$ (and, consequently, $\sigma(B_{m_j}^{(n_j)}) = 2^{n_j} - p^{m_j}$), where the sets $A_{m_j}^{(n_j)}$ are constructed by the following procedure. We set $A_{m_1}^{(n_1)} = \{x_1, \ldots, x_{p^{m_1}}\}$, where $X^{n_1} = \{x_1, \ldots, x_{p^{m_1}}, \ldots, x_{2^{n_1}}\}$. Suppose that the set $A_{m_j}^{(n_j)}$ has been constructed. We set

$$C_{m_{j+1}}^{n_{j+1}} = \{x \in X^{n_j} : x \text{ has a prefix } y \text{ belonging to the set } B_{m_j}^{(n_j)} = X^{n_j} \backslash A_{m_j}^{(n_j)}\}$$

$$= B_{m_j}^{(n_j)} \times X^{n_{j+1} - n_j}.$$

The set $A_{m_{j+1}}^{(n_{j+1})}$ is the union of the set $C_{m_{j+1}}^{n_{j+1}}$ and $p^{m_{j+1}} - \sigma(C_{m_{j+1}}^{n_{j+1}}) = p^{m_{j+1}} - N_j$ (where $N_j = (2^{n_j} - p^{m_j})2^{n_{j+1} - n_j}$) first elements of $X^{n_{j+1}} = (x_1, \ldots, x_{2^{n_{j+1}}})$ which do not belong to $C_{m_{j+1}}^{n_{j+1}}$. By (7.2) $p^{m_{j+1}} - N_j > 0$ for all $j \geq 1$. Thus this procedure define sets $A_{m_j}^{(n_j)}$ for all $j \geq 1$.

We set $V_{m_k} = \cup_{j=k}^{\infty} A_{m_j}^{(n_j)}$ and $V_m = V_{m_k}$ for $m_{k-1} < m \leq m_k$. We prove that $V = \{(x, m) : x \in V_m\}$ is a p-adic test. The set V is r.e. and $V_m \supset V_{m+1}$ by the procedure of construction. We also have: $V_{m_k}^{(n_j)} = A_{m_j}^{(n_j)}$, $j \geq k$, and $V_{m_k}^{(n)} = \emptyset$, $n \neq n_j$, $j \geq k$. Thus

$$|\sigma(V_{m_k}^{(n_j)})|_p = |\sigma(A_{m_j}^{(n_j)})|_p = p^{-m_j} \leq p^{-m_k}.$$

On the other hand, for each m_k, we have:

$$(\cup\{\mathcal{U}_x : x \in A_{m_k}^{(n_k)}\}) \cup (\cup\{\mathcal{U}_x : x \in A_{m_{k+1}}^{(n_{k+1})}\})$$

$$\supset (\cup\{\mathcal{U}_x : x \in A_{m_k}^{(n_k)}\}) \cup (\cup\{\mathcal{U}_{ya} : y \in B_{m_k}^{n_k}, a \in X^{n_{k+1} - n_k}\})$$

$$= (\cup\{\mathcal{U}_x : x \in A_{m_k}^{(n_k)}\}) \cup (\cup\{\mathcal{U}_y : y \in B_{m_k}^{n_k}\}) = X^\infty.$$

■

The previous result imply that in the p-adic case (similar to Schnorr's theory of randomness [98]) the only reasonable approach to randomness of infinite sequences is to use randomness with respect to the concrete p-adic sequential test O. Of course, the use of O-randomness has extremely different origins in our theory and Schnorr's theory. It seems that in the p-adic case this situation is a consequence of the impossibility to define σ-additive (non-discrete) probability on the σ-algebra generated by $\mathcal{F}_{cyl}$. Thus we have no other possibility than to identify a sequential tests O with tests $V \subset X^* \times \mathbf{N}$. In Schnorr's theory this situation is a consequence of the use of total recursive null-sets.

Bibliography

1. Accardi L., The probabilistic roots of the quantum mechanical paradoxes. *The wave-particle dualism. A tribute to Louis de Broglie on his 90th Birthday*, Edited by S. Diner, D. Fargue, G. Lochak and F. Selleri. D. Reidel Publ. Company, Dordrecht, 47–55 (1970).

2. Accardi L., *Urne e Camaleoni: Dialogo sulla realta, le leggi del caso e la teoria quantistica.* Il Saggiatore, Rome (1997).

3. Albeverio S., Fenstad I., Hoegh-Krohn P., and Lindström T., *Nonstandard Method in the Stochastic Analysis and Mathematical Physics.* Academic Press, London,... (1990).

4. Albeverio S., Khrennikov A.Yu., Representation of the Weyl group in spaces of square integrable functions with respect to p-adic valued Gaussian distributions. *J. of Phys. A*, *29*, 5515–5527 (1996).

5. Aspect A., Dalibard J., Roger G., Experimental test of Bell's inequalities using time-varying analyzers. *Phys. Rev. Lett.*, **49**, 1804-1807 (1982).

6. Atmanspacher H., Räth C., Wiedenmann G., Statistics and metastatistics in the concept of complexity. *Physica A*, **234**, 819-829 (1997).

7. Ballentine L.E., *Quantum mechanics.* Englewood Cliffs, New Jersey (1989).

8. Ballentine L.E., The statistical interpretation of quantum mechanics. *Rev. Mod. Phys.*, **42**, 358–381 (1970).

9. Barnett M.S., Negative probability. *Proc. Cambridge Phil. Soc.*, **41**, 71-73 (1944).

10. Bell J.S, On the Einstein–Podolsky–Rosen paradox. *Physics*, **1**, 195-200 (1964).

11. Bell J.S., On the problem of hidden variables in quantum mechanics. *Rev. Mod. Phys.*, **38**, 447–452 (1966).

12. Bell J., *Speakable and unspeakable in quantum mechanics.* Cambridge Univ.Press (1987).

13. Bohm D. and Hiley B., *The undivided universe: an ontological interpretation of quantum mechanics.* Routledge and Kegan Paul, London (1993).

14. Bohm D., Quantum theory as an indications of a new order in physics. Part B. Implicate and explicate order in physical law. *Found. Phys.*, **1**, 139-168(1971).

15. Borel E., *Éléments de la théorie des probabilités.* Paris (1908).

16. Borel E., Les probabilités dénombrables et leurs applications arithmétiques. *Rend. del Circolo matem. di Palermo* (1990).

17. Buonomano V., *Quantum uncertainties, Recent and Future Experiments and Interpretations*, edited by W.M. Honig, D.W. Kraft and E. Panarella, NATO ASI Series, **162**, Plenum Press, New York (1986).

18. Chaitin G. J. , On the length of programs for computing finite binary sequences. *J. Assoc. Comput. Mach.*, **13**, 547- 569 (1966).

19. Church A., On the concept of a random sequence. *Bull. Amer. Math. Soc.*, **46**, 130-135 (1940).

20. Clauser J.F, Shimony A., Bell's theorem: experimental tests and implications. *Rep. Progr.Phys.*, **41** 1881-1901 (1978).

21. Clauser J.F., Horne M.A., Schimony A. and Holt R.A., *Phys. Rev. Lett.*, **23**, 888–893 (1969).

22. Clauser J.F. and Horne M.A., *Phys. Rev. D.*, **10**,526–547 (1974).

23. Cramer H., *Mathematical theory of statistics.* Univ. Press, Princeton (1949).

24. Davies E. B., *Quantum theory of open systems*, Academic Press, London (1976).

25. De Baere W., On the significance of Bell's inequality for hidden-variable theories. *Lett. Nuovo Cimento*, **39**, 234-238 (1984).

26. De Baere W., Einstein–Podolsky–Rosen paradox. *Adv. Electronics and Electron physics*, **68**, 245-336.

27. De Broglie L., *The current interpretation of wave mechanics. A critical study*, Elsevier Publ. co., Amsterdam (1964).

28. De Broglie L., *La Physique Quantique Restera–t–elle Indeterministe?* Gauthier-Villars, Paris (1953).

29. d'Espagnat B., *Reality and Physicist.* Cambridge Univ. Press (1989).

30. d'Espagnat B., *Veiled Reality. An anlysis of present-day quantum mechanical concepts.* Addison-Wesley publishing company (1995).

31. De Muynck W., Can we escape from Bell's conclusion that quantum mechanics describes a non-local reality. *Stud. Hist. Phil. Mod. Phys.*, **27**, N.3, 315-330 (1996).

32. De Muynck W. and De Baere W., Quantum nonlocality and Bell's inequality. *Ann. Israel Phys. Soc.*, **12**, 1-22 (1996).

33. De Muynck W.M., De Baere W., Martens H., Interpretations of quantum mechanics, joint measurement of incompatible observables, and counterfactual definitness. *Found. of Physics*, **24**, 1589–1663 (1994).

34. Dirac P.A.M., The physical interpretation of quantum mechanics. *Proc. Roy. Soc. London*, **A 180**, 1–39 (1942).

35. Dubischar D., Gundlach V.M., Steinkamp O., Khrennikov A. Yu., A p-adic model for the process of thinking disturbed by physiological and information noise. *J. Theor. Biology,* **197**, 451-467 (1999).

36. Einstein A., Podolsky B., Rosen N., Can quantum mechanical description of physical reality be considered complete? *Phys. Rev.*, **47**, 777–780 (1935).

37. Feynman R.P., Negative probability. *Quantum Implications, Essays in Honour of David Bohm*, B.J. Hiley and F.D. Peat, editors. Routledge and Kegan Paul, London, 235–246 (1987).

38. Fine A., Hidden variables, joint probability, and Bell inequalities. *Phys. Rev. Letters*, **48**, 291–295 (1982).

39. Fine T.L., *Theories of probabilities, an examination of foundations.* Academic Press, New York (1973).

40. Frechet M., *Recherches théoriques modernes sur la théorie des probabilités.* Univ. Press., Paris (1937-1938).

41. Freund P. G. O. and Witten E., Adelic string amplitudes.*Phys. Lett. B* , **199**, 191–195 (1987).

42. Gudder S.P., Probability manifolds. *J. Math Phys.,* **25**, 2397- 2401 (1984).

43. Heisenberg W., *Physics and Philosophy.* Allen and Unwin, London (1958).

44. Hensel K., Untersuchung der Fundamentalgleichung einer Gattung für eine reelle Primzahl als Modul und Bestimmung der Theiler ihrer Discriminante. *J. Reine Angew Math.*, **113**, 61–83 (1894).

45. Hida T., *Brownian Motion.* Nauka, Moscow (1987).

46. Holevo A. S., *Probabilistic and statistical aspects of quantum theory.* North-Holland, Amsterdam (1982).

47. Home D., Selleri F., *Nuovo Cim. Rivista,* **14**, 2–176 (1991).

48. Jech T.J., *The axiom of choice.* North-Holland publishing company, Amsterdam (1973).

49. Kamke E., Über neuere Begründungen der Wahrscheinlichkeitsrechnung. *Jahresbericht der Deutschen Matematiker-Vereinigung,* **42,** 14-27 (1932).

50. Keynes J.M., *A treatise on Probability.* Macmillan, London (1952).

51. Khrennikov A.Yu., Integration with respect to generalized measures on linear topological spaces. *Trudy Mosk.Mat.Obshch.,* **49**, 113–129 (1986).

52. Khrennikov A.Yu., *Supernalysis.* Nauka, Moscow, (1997), in Russian. English translation: Kluwer Academic Publishers (1999).

53. Khrennikov A.Yu., p-adic probability predictions of correlations between particles in the two slit and neuron interferometry experiments. *Il Nuovo Cimento B,* **113,** N.6, 751-760 (1998).

54. Khrennikov A.Yu., Description of experiments detecting p-adic statistics in quantum diffraction experiments. *Doklady Mathematics (Doklady of Russian Academy of Sc.),* **58,** N. 3, 478-480 (1998).

55. Khrennikov A. Yu., Mathematical methods of the non-Archimedean physics. *Uspekhi Mat. Nauk*, **45**, No. 4, 79–110 (1990).

56. Khrennikov A. Yu., The p-adic probability and statistics. *Dokl. Akad. Nauk SSSR, ser. Matem.*, **322**, No. 6, 1075–1079 (1992).

57. Khrennikov A. Yu., p-adic quantum mechanics with p-adic valued functions. *J. Math. Phys.*, **32**, No. 4, 932–937 (1991).

58. Khrennikov A.Yu., p-adic probability interpretation of Bell's inequality paradoxes. *Physics Letters A*, **200**, 119–223 (1995).

59. Khrennikov A.Yu., p-adic probability distribution of hidden variables. *Physica A*, **215**, 577–587 (1995).

60. Khrennikov A.Yu., *p-adic valued distributions in mathematical physics.* Kluwer Academic Publishers, Dordrecht (1994).

61. Khrennikov A.Yu., p-adic probability description of Dirac's hypothetical world. *Int. J. Theor. Phys.*, **34**, No.12, 2423–2434 (1995).

62. Khrennikov A.Yu., Statistical interpretation of p-adic quantum theories with p-adic valued wave functions. *J. Math. Phys.*, **36**, No.12, 6625–6632 (1995).

63. Khrennikov A.Yu., Ultrametric Hilbert space representation of quantum mechanics with a finite exactness. *Found. of Physics.*, **26**, No. 8, 1033–1054 (1996).

64. Khrennikov A.Yu., *Time to change Kolmogorov axiomatics*, Preprint Ruhr University Bochum, SFB-237, No. 297 (1996).

65. Khrennikov A.Yu., *Non-Archimedean analysis: quantum paradoxes, dynamical systems and biological models.* Kluwer Acad.Publ., Dordreht, The Netherlands (1997).

66. Khrennikov A.Yu., Human subconscious as the p-adic dynamical system. *J. of Theor. Biology,* **193,** 179-196 (1998).

67. Khrennikov A.Yu., p-adic stochastic hidden variable model. *J. of Math. Physics,* **39,** No. 3, 1388-1402 (1998).

68. Khrennikov A.Yu., *Classical and quantum mechanics on information spaces with applications to cognitive, psychological, social and anomalous phenomena.* Växjö University, MASDA, Report 9824; *Found. of Physics,* **29**, N. 7 (2000).

69. Khrennikov A.Yu., *p-adic information dynamics and the pilot wave theory for cognitive systems.* Växjö University, MASDA, Report 9839.

70. Khrennikov A.Yu., p-adic description of chaos. *Proc. of Workshop "Nonlinear Physics:theory and experiment"*, Gallipoli,Italy, 1995. Editors E.Alfinito, M.Boti,..., World Sci. Publ., Singapure, 177–184 (1996).

71. Khrennikov A.Yu., *The interference phenomenon and random dynamical systems over the fields of p-adic numbers.* Växjö University, MASDA, Report 9810.

72. Khrennikov A.Yu., Yamada Sh., *Tests for randomness for p-adic probability theory.* Växjö University, MASDA, Report 9915.

73. Khrennikov A.Yu., van Rooij A., Yamada Sh., The measure-theoretical approach to p-adic probability theory. *Annales Math. Blaise Pascal,* N.1 (2000).

74. Kolmogoroff A. N., *Grundbegriffe der Wahrscheinlichkeitsrechnung.* Springer Verlag, Berlin (1933); reprinted: *Foundations of the Probability Theory.* Chelsea Publ. Comp., New York (1956).

75. Kolmorogof A.N., Über die analytischen Methoden in der Wahrschein-lichtkeitsrechnung. *Math. Ann.*, **104**, 415–458 (1931).

76. Kolmolgorov A. N., Logical basis for information theory and probability theory. *IEEE Trans.*,**IT–14**, 662–664 (1968).

77. Kolmolgorov A. N. , Three approaches to the quantitative definition of information. *Problems Inform. Transmition.*, **1**, 1-7 (1965).

78. Kolmorogov A.N., Fomin S.V., *Elements of Theory of Functions and Functional Analysis.* Nauka, Moscow (1980).

79. Li M., Vitànyi P., *An introduction to Komogorov complexity and its applications.* Springer, Berlin-Heidelberg-New York (1997).

80. Lochak G., De Broglie's initial conception of de Broglie waves. *The wave–particle dualism. A tribute to Louis de Broglie on his 90th Birthday,* Edited by S. Diner, D. Fargue, G. Lochak and F. Selleri. D. Reidel Publ. Company, Dordrecht, 1–25 (1970).

81. Ludwig G., *Foundations of quantum mechanics.* Springer Verlag, Berlin (1983).

82. Maddox J., Can chance be less than zero? *Nature,* **320**, 10 April (1986).

83. Martin-Löf P., On the concept of random sequence. *Theory of Probability Appl.*, **11**, 177-179(1966).

84. Martin-Löf P., The definition of random sequence. *Inform. Contr.,* **9**, 602-619 (1966).

85. Martin-Löf P., Complexity oscillations in infinite binary sequences. *Z. Wahrscheinlichkeitstheorie verw.,* **19**, 225-230 (1971).

86. Mises R. von, Grundlagen der Wahrscheinlichkeitsrechnung. *Math.Z.,* **5**, 52–99 (1919).

87. Mises R. von, *Probability, Statistics and Truth.* Macmillan, London (1957).

88. Mises R. von, *The mathematical theory of probability and statistics.* Academic, London (1964).

89. Monna A. and Springer T., Integration non-Archimédienne, 1,2. *Indag. Math.*, **25**, 634–653 (1963).

90. Muckenheim W., A review on extended probabilities. *Phys. Reports,* **133**, 338–401 (1986).

91. Obata N., Density of natural numbers and the Levy group. *J. of Number Theory,* **30**, 288–297(1988).

92. Pitowsky I., Resolution of the Einstein–Podolsky–Rosen and Bell paradoxes. *Phys. Rev. Lett,* **48**, N.10, 1299-1302 (1982).

93. Pitowsky I., Deterministic model of spin and statistics. *Phys. Rev. D,* **27**, N.10, 2316-2326 (1983).

94. Prugovecki E., Simultaneous measurement of several observables. *Found. of Physics,* **3**, No. 1, 3–18 (1973).

95. Rastall P., The Bell inequalities. *Found. of Physics,* **13**, 555-570 (1983).

96. Rauch H., Summhammer J., Zawisky M., Jericha E., Law–contrast and low-counting-rate measurements in neutron interferometry. *Phys. Rev. A*, **42**, 3726–3732 (1990).

97. Schikhov W., *Ultrametric calculus.* Cambridge Univ. Press, Cambridge (1984)

98. Schnorr C. P., Zufälligkeit und Wahrscheinlichkeit. *Lect. Notes in Math.,* **218**, Springer-Verlag, Berlin (1971).

99. *Selected works of A.N. Kolmogorov,* **2,** *Probability theory and mathematical statistics.* Ed. A. N. Shiryayev. Kluwer Academic Publ., Dordrecht (1986).

100. Solomonoff R. J., A formal theory of inductive inference. *Inform. Contr.,* **7**, 1-22, 224-254 (1964).

101. Summhammer J., Neutron interferometric test of the nonergodic interpretation of quantum mechanics. *Il Nuovo Cimento*, **103 B**, 265–280 (1989).

102. Tornier E., *Wahrscheinlichkeitsrechnunug und allgemeine Integrationstheorie.*, Univ. Press, Leipzing (1936).

103. Van Lambalgen M., Von Mises' definition of random sequences reconsidered. *J. of Symbolic Logic,* **52,** N. 3 (1987).

104. Van Rooij A., *Non-archimedean functional analysis.* Marcel Dekker, Inc., New York (1978).

105. Venn J., *The logic of chance.* London (1866); reprint: Chelsea, New York (1962).

106. Ville J., *Etude critique de la notion de collective.* Gauthier– Villars, Paris (1939).

107. Vladimirov V. S., Volovich I. V., and Zelenov E. I., *p-adic analysis and Mathematical Physics.* World Scientific Publ., Singapure (1993).

108. Volovich I. V., p-adic string. *Class. Quant. Grav.*, **4**, 83–87 (1987).

109. von Neuman J., *Mathematical foundations of quantum mechanics.* Princeton Univ. Press, Princenton (1955).

110. Wald A., Die Widerspruchsfreiheit des Kollektivbegriffs in der Wahrscheinlichkeitsrechnung. *Ergebnisse eines Math. Kolloquiums*, **8**, 38-72 (1938).

111. Wigner E., *Quantum -mechanical distribution functions revisted,* in: *Perspectives in quantum theory.* Yourgrau W. and van der Merwe A., editors, MIT Press, Cambridge MA (1971).

112. Zvonkin A. K. and Levin L. A., A complexity of finite objects and justification of the notions of information and randomness on the basis of algorithms. *Uspehi Matem. Nauk*, **25**, No. 6, 85–127 (1970).

Index

Algebra of sets 13
Bayes' formula 16
Bell inequalities 89, 94
Bell mystification 92
Cardinality 14
Characteristic function 14
Cognitive system
pilot wave theory 185
consciousness 186
Collective 18
combining 48
independent 51
mixing 20
partition 20
Density of natural numbers 26
Dirac's negative probability 135
Disturbance effect 68
Ergodicity 38
Gumbling strategy 19
Formula of total probability 16
Hidden variables 95, 125, 127
deteministic 96
stochastic 100
Information
active 176
classical 173
energy 174
field 175

mass 177
motivation 174
quantum 180
state 174
time 173
Interference 75
Interpretation of probability
classical 14
ensemble 15
frequency 18
ensemble-frequency (conventional) 32
Interpretation of wave function
ensemble realists 64
ensemble empirisists 65
individual empirisists 65
individual realists 65
nonergodic 77
orthodox Copenhagen 65
probability 62
Independence 23
Kolmogorov 22
complexity 45
axiomatics 22
Law of large numbers 36
Measure 13
p-adic 163
p-adic uniform 196
signed 120
Measurable 13
Nonreproducibility
statistical 99, 104
individual 105
Non-Kolmogorovean models
Accardi 106
Pitowsky 106
Gudder 107
Numbers
p-adic 147

m-adic 147
Objection
Kamke 41
Ville 44
Paradox
Banach–Tarski 25
Einstein–Podolsky–Rosen 83
Place selection 18
Bernoulli 42
Church 43
Probability
classical 14
comparative 56
conditional 15, 21, 23
ensemble 15
fluctuating 82, 109
frequency 17
negative 109, 119, 124, 125
nonuniqueness 30
p-adic 149, 152
space 22
subjective 39
uniform 14
Property 59
empirisism 60
idealism 61
realism 60
Procedures
preparation 66
measurement 66
Randomness 18
Random variable 22
Reality
element 84
information 171
Newton 171
nonlocal 85
Recursive enumeration 205

Semi-
  algebra 27
  measure 27
  probability 27
Statistical stabilization 18
Strong triangle axiom 144
Test for randomness
  Martin-Löf 45
  p-adic 199
  Schnorr 45
Theorem
  Bernoulli 36
  Bayes 40
  Hahn–Banach 29
  Hahn–Jordan 121
  limit 203
  Wald 43
Ultrametric 145
Valuation 144
  Archimedean 144
  non-Archimedean 144
von Mises 17
Wiegner distribution 128
  harmonic oscillator 131